

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:50:22 ; Search time 53.04 seconds
(without alignments)
1614.489 Million cell updates/sec

Title: US-09-722-544A-2MOD
Perfect score: 2632
Sequence: 1 MKIKAEKNEGPRSRWQLHW.....FIGYVAKALQHKEHFKTW 495

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2460	93.5	466	Q15846	Q15846 homo sapien
2	1832.5	69.6	465	Q95KN1	Q95KN1 canis famil
3	1226.5	46.6	338	Q9NIT8	Q9NIT8 canis famil
4	423	16.1	449	Q29482	Q29482 equus cabal
5	402.5	15.3	448	Q9VGP0	Q9VGP0 gallus gall
6	321	12.2	372	Q9JK98	Q9JK98 mus musculu
7	266.5	10.1	295	Q35510	Q35510 rattus norv
8	204	7.8	218	Q15846	Q15846 homo sapien
9	164	6.2	1738	Q9VGP0	Q9VGP0 gallus gall
10	157.5	6.0	3616	Q9VGP0	Q9VGP0 gallus gall
11	150	5.7	1156	Q9VGP0	Q9VGP0 gallus gall
12	149.5	5.7	840	Q9VGP0	Q9VGP0 gallus gall
13	148.5	5.7	1388	Q9VGP0	Q9VGP0 gallus gall
14	148.5	5.6	1091	Q18082	Q18082 caenorhabdi
15	147	5.6	3259	Q14789	Q14789 homo sapien
16	145	5.5	1388	Q075116	Q075116 homo sapien

17	145	5.5	1388	4	Q9UQN5	Q9UQN5 homo sapien
18	144	5.5	1379	11	Q62868	Q62868 rattus norv
19	143	5.4	1931	5	Q9NCF9	Q9NCF9 drosophila
20	143	5.4	1931	5	Q9VKH9	Q9VKH9 drosophila
21	142.5	5.4	533	6	Q95JY2	Q95JY2 macaca fasc
22	140.5	5.3	2166	16	O51465	O51465 borrelia bu
23	140	5.3	1132	4	Q75065	Q75065 homo sapien
24	140	5.3	1305	10	Q9FJ35	Q9FJ35 arabidopsis
25	138.5	5.3	1391	11	Q922J3	Q922J3 mus musculu
26	138	5.2	1084	16	O83423	O83423 treponema p
27	138	5.2	1955	5	O61308	O61308 parascaris
28	136.5	5.2	684	3	Q07238	Q07238 pneumocysti
29	136.5	5.2	709	5	O9GRG1	O9GRG1 tetrahymena
30	136.5	5.2	720	4	Q9H6Q7	Q9H6Q7 homo sapien
31	136.5	5.2	1410	4	Q14221	Q14221 homo sapien
32	136	5.2	1388	11	P70336	P70336 mus musculu
33	135.5	5.1	1956	5	Q20641	Q20641 caenorhabdi
34	135	5.1	1133	5	Q21022	Q21022 caenorhabdi
35	135	5.1	1330	6	O97961	O97961 vulpes vulp
36	135	5.1	1530	4	O43241	O43241 homo sapien
37	135	5.1	5458	5	Q9U459	Q9U459 plasmodium
38	134.5	5.1	1300	4	Q13999	Q13999 homo sapien
39	134.5	5.1	2035	11	Q9JMH9	Q9JMH9 mus musculu
40	134	5.1	1538	13	Q98940	Q98940 gallus gall
41	133.5	5.1	2138	5	O9XZE3	O9XZE3 amoeba prot
42	133	5.1	1938	6	O9GJF9	O9GJF9 oryctolagus
43	132.5	5.0	1411	4	Q15075	Q15075 homo sapien
44	132.5	5.0	1708	5	Q9U0S6	Q9U0S6 mytilus gal
45	132.5	5.0	2473	11	Q9QZ84	Q9QZ84 mus musculu

ALIGNMENTS

RESULT 1
ID Q15846 PRELIMINARY; PRT; 466 AA.
AC Q15846;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 54.2 KDA PROTEIN PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shimizu A., Nishida K., Kinoshita G., Imazawa I., Okubo K.,
RA Matsubara K.,
RT "Expression profile of active genes in human retina."
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; D63813; BAA09882.1; -
DR Interpro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 3.
DR SMART; SM00035; Clb; 1.
DR SMART; SM00030; Clb; 1.
KW Signal; Hypothetical protein.
FT SIGNAL 1 20
FT CHAIN 21 466
SQ SEQUENCE 466 AA; 54214 MW; 41A603D0FD6439C4 CRC64;
POTENTIAL.
ROD PHOTORECEPTOR PROTEIN.

Query Match 93.5%; Score 2460; DB 4; Length 466;
Best Local Similarity 99.3%; Pred. No. 1.2e-160;
Matches 465; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 30 MKPPLLVFVCLLWLDKSCAPTWKDKTAISENLKSFSEVGEIDAEVKKALTGKQMK 89
Db 1 MKPPLLVFVCLLWLDKSCAPTWKDKTAISENLKSFSEVGEIDAEVKKALTGKQMK 60
Qy 90 IMMERKEKEHTNLMSTLKKCKREEKQEAALKLLNEVQHLLEERLCRESLADSGWGCRCSL 149

Sequence
to diff
than sea IDW.

Does
not display
HENG
gap product
- is not linked
to Schizophrina

Db	61	IMMERKEHTNLMSTLKKREKQKALKNLNEVOEHLEEBERLCRESLADSWGCRSCL	120
Qy	150	ENNCRIYTTCCPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQM	209
Db	121	ENNCRIYTTCCPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQM	180
Qy	210	EDVFSQLTVDVNSLNRSENFVFRMQQEFQDTQFSHFISDTDLTPYFFPAFSKEPMTKA	269
Db	181	EDVFSQLTVDVNSLNRSENFVFRMQQEFQDTQFSHFISDTDLTPYFFPAFSKEPMTKA	240
Qy	270	DLEQCWDIPNFFQLCNFSVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ	329
Db	241	DLEQCWDIPNFFQLCNFSVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ	300
Qy	330	DRGLCGELDONLSRCFKHEKQKCOAHLSEDCPDVPAHLTELDEAIRLVNVSNOOYGOI	389
Db	301	DRGLCGELDONLSRCFKHEKQKCOAHLSEDCPDVPAHLTELDEAIRLVNVSNOOYGOI	360
Qy	390	LQMKRKHLEDLAYLVEKMRGQFGWVSELANQAPETETIFNSIQVVPRIHEGNSKQDET	449
Db	361	LQMKRKHLEDLAYLVEKMRGQFGWVSELANQAPETETIFNSIQVVPRIHEGNSKQDET	420
Qy	450	MTDLSILPSSNFTLKIPLESSESSNFIYGVYVAKALQHEKHEFTW	495
Db	421	MTDLSILPSSNFTLKIPLESSESSNFIYGVYVAKALQHEKHEFTW	466
RESULT	2		
Q95KNI			
ID	Q95KNI	PRELIMINARY;	PRT; 465 AA.
AC	Q95KNI		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	RETINAL CLUSTERIN-LIKE PROTEIN CLULB SPICE VARIANT.		
OS	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_TaxID=9615;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=RETINA;		
RC	MEDLINE=20156379; PubMed=10675623;		
RA	Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Aguirre G.D.;		
RT	"Molecular cloning, characterization and expression of a novel retinal		
RT	clusterin-like protein cDNA."		
RL	Gene 243:151-160(2000).		
DR	EMBL; AF241221; AAK43030.1; --		
SQ	SEQUENCE 465 AA; 54402 MW; BBDELAC512D5D33F CRC64;		
Query Match	69.6%;	Score 1832.5;	DB 6; Length 465;
Best Local Similarity	73.2%;	Pred. No. 1.1e-117;	
Matches	341;	Conservative 52;	Mismatches 72; Indels 1; Gaps 1;
Qy	30	MKPPLLVFVCLLWLDKSHCAPTKWOKTAISENLKSFSEVGEIDAEVKKALGTIKQWK	89
Db	1	MKPSLLVFTVYLLWLDKSHCAPTKWOKTDHGNLKGFESEAGDIDVDEEVKKALIGNKQWK	60
Qy	90	IMMERKEHTNLMSTLKKREKQKALKNLNEVOEHLEEBERLCRESLADSWGCRSCL	149
Db	61	IMMERKEHTNLMSTLKKREKQKALKNLNEVOEHLEEBERLCRESLADSWGCRSCL	120
Qy	150	ENNCRIYTTCCPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQM	209
Db	121	ENNCRIYTTCCPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQM	180
Qy	210	EDVFSQLTVDVNSLNRSENFVFRMQQEFQDTQFSHFISDTDLTPYFFPAFSKEPMTKA	269
Db	181	EDVFSQLTVDVNSLNRSENFVFRMQQEFQDTQFSHFISDTDLTPYFFPAFSKEPMTKA	240
Qy	270	DLEQCWDIPNFFQLCNFSVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ	329

Db	241	DPVQSWDIPSFQFLYFNFSLSIYHSITTTITKTLNAIEDLPKQDNDSNHGSLSKTLPVQ	300
Qy	330	DRGLCGELDONLSRCFKHEKQKCOAHLSEDCPDVPAHLTELDEAIRLVNVSNOOYGOI	389
Db	301	DRGLCGELDONLSRCFKHEKQKCOAHLSEDCPDVPAHLTELDEAIRLVNVSNOOYGOI	360
Qy	390	LQMKRKHLEDLAYLVEKMRGQFGWVSELANQAPETETIFNSIQVVPRIHEGNSKQDET	449
Db	361	LQMKRKHLEDLAYLVEKMRGQFGWVSELANQAPETETIFNSIQVVPRIHEGNSKQDET	420
Qy	450	MTDLSILPSSNFTLKIPLESSESSNFIYGVYVAKALQHEKHEFTW	495
Db	421	MTDLSILPSSNFTLKIPLESSESSNFIYGVYVAKALQHEKHEFTW	465
RESULT	3		
Q9NIT8			
ID	Q9NIT8	PRELIMINARY;	PRT; 338 AA.
AC	Q9NIT8		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	RETINAL-SPECIFIC CLUSTERIN-LIKE PREPROTEIN.		
GN	CLUL1.		
OS	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_TaxID=9615;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	MEDLINE=20156379; PubMed=10675623;		
RA	Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Aguirre G.D.;		
RT	"Molecular cloning, characterization and expression of a novel retinal		
RT	clusterin-like protein cDNA."		
RL	Gene 243:151-160(2000).		
DR	EMBL; AF147784; AAF36799.1; --		
DR	InterPro; IPR000753; Clusterin.		
DR	SMART; SM00035; Cla; 1.		
DR	SMART; SM00030; CLB; 1.		
SQ	SEQUENCE 338 AA; 39385 MW; B86F07877213FCC2 CRC64;		
Query Match	46.6%;	Score 1226.5;	DB 6; Length 338;
Best Local Similarity	69.8%;	Pred. No. 2.2e-76;	
Matches	231;	Conservative 42;	Mismatches 57; Indels 1; Gaps 1;
Qy	165	SSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQMEDVFSQLTVDVNSLF	224
Db	9	TSGTSTVEQFFRNIIYQFLFPFHEDNEKDLPISEKLIKDAQLTQMEDVFSQLTVDVNSLF	68
Qy	225	NRSFNVRMQQEFQDTQFSHFISDTDLTPYFFPAFSKEPMTKADLCQWDIPNFFQF	284
Db	69	NRSFNVRMQQEFQDTQFSHFISDTDLTPYFFPAFSKEPMTKADLCQWDIPNFFQF	128
Qy	285	CNFSVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGDRLCGELDONLSRC	344
Db	129	YNFSLSIYHSITTTITKTLNAIEDLPKQDNDSNHGSLSKTLPVQHRPGYGEFGNLSEC	188
Qy	345	FRFHEKQKCOAHLSEDCPDVPAHLTELDEAIRLVNVSNOOYGOIQLMKRKHLEDLAYL	404
Db	189	FQFHARCOQCDYLWEDCPDVPHELTKVDEALELYNISHQQAQVQLQMTQHLEDITVLM	248
Qy	405	EKMRGQFGWVSELANQAPETETIFNSIQVVPRIHEGNSKQDETMTDLSILPSSNFTLK	464
Db	249	EKMRGQFGWVSELANQAPETETIFNSIQVVPRIHEGNSKQDETMTDLSILPSSNFTLK	307
Qy	465	IPLESSESSNFIYGVYVAKALQHEKHEFTW	495
Db	308	IPLESSESSNFIYGVYVAKALQHEKHEFTW	338
RESULT	4		

```
Q29482
ID Q29482 PRELIMINARY; PRT; 449 AA.
AC Q29482;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CLUSTERIN PRECURSOR.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OLIVER; TISSUE=TESTIS;
RA Barber J.A., Farris J.A., Troedsson M.H.T., Crabo B.G., Foster D.;
RT "Nucleotide sequence of the complementary DNA encoding Equine
Clusterin."
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR EMBL; L46797; AA080313.1; -.
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; CLA; 1.
DR SMART; SM00030; CLB; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 449 CLUSTERIN.
SQ SEQUENCE 449 AA; 52154 MW; 705A60504515F7C0 CRC64;

Query Match 16.1%; Score 423; DB 6; Length 449;
Best Local Similarity 24.5%; Pred. No. 3.2e-21;
Matches 117; Conservative 100; Mismatches 203; Indels 58; Gaps 13;

QY 34 LLVFTVCLLLKDSHCAPTWDKTAISENLKSFSEVEIDADEVKALGIKQMKIME 93
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4 LLLVGLLLTLENGQ---VLGDKAVSDRELGQEMSTQGSNTINKALGVKQIKNLIE 60

QY 94 RKEHTNLMSTLKKREEQKALLLNEVOHLEERLCRESLADSWGECRSCLENNC 153
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 QTNEERKSLGLTLEAKKKEGALNDTKDSMKLKESQGVNETMTALWEECKPCLKOTC 120

QY 154 MRYIT-TCOPSSWSSVKNKIERFRKIYQFLPFPHEDNEKOLPISEKLIKDAOLTMEDY 212
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
121 MKFYARVCRSGSLVGHQLEEFLLNQSSPFYFWINGDR-----IDSLLENDROQTHVLDV 174

QY 213 FSQLTVDVNSLFNRSFNVFMQOQEFQDTQSFHSFISDPTLPEYFPFAFSKEPMTKADL- 271
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
175 -----MQDSFDRASSI-----MDELQDRFFT-REPQDIYYSPFS-SPHRSSLL 218

QY 272 -----EQCWDIP-----NFTOLFNFVSIVYESYSETITKMLKAIEDLPKQD--KAP 316
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
219 FNPKSREARNIMHPMYRHILNFMDFPFDMTHQAQAMNLHLRLPDQLPMTSESGD 278

QY 317 DHGGLISKMLPGDGLCGELDONLRCFKFHEKCKQCOAHLSEDC-----PDVPAHLTEL 372
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
279 NH-----DRTVCKEIRHNSTGLKMKDOCEKQCQELISVDCSTNNPNSQMQLRQEL 327

QY 373 DEAIRLVNSNOQYGOILOMTRKHELTAYLVEKMRGQFGVWSELANOAPETEIFNSIQ 432
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
328 NNSLQAEKFTKLYDELQSYQEKMLNTSSLLKQLNEQFSWVSQLANLTQGEDQYYLYQVT 387

QY 433 VVPRIHGNTSKQDETMWTLISPLSPNSFTLKIPIEESAESSNFIGYVVAKALQHFKE 490
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
388 TVSS-HNSDSEVPFSGLRVVVVKLFDSYPIVTVPVSVSRNPNKFMETVAEKALQEYRQ 444
```

```
RESULT 5
Q9YGP0
ID Q9YGP0 PRELIMINARY; PRT; 448 AA.
AC Q9YGP0;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CLUSTERIN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99134326; PubMed-9933599;
RA Mahon M.G., Lindstedt K.A., Hermann M., Nimpf J., Schneider W.J.;
RT "Multiple involvement of clusterin in chicken ovarian follicle
development. Binding to two oocyte-specific members of the low density
lipoprotein receptor gene family."
RL J. Biol. Chem. 274:4036-4044(1999).
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR EMBL; AF119370; AAD17257.1; -.
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; CLA; 1.
DR SMART; SM00030; CLB; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
KW Glycoprotein.
SQ SEQUENCE 448 AA; 51348 MW; 9982873DCA13D27C CRC64;

Query Match 15.3%; Score 402.5; DB 13; Length 448;
Best Local Similarity 26.0%; Pred. No. 8.1e-20;
Matches 118; Conservative 87; Mismatches 189; Indels 59; Gaps 14;

QY 63 LKSFSEVGIEDAEEVKKALTGKQKIMMERKEHTNLMSTLKKREEQKALKLINE 122
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
27 LKOLSAAGSKYIDTEVENAINGVKQKMTLMDKTSKHEQAMLHLEETKRKEEAVKLALE 86

QY 123 VOHLEEEERLCRESLADSWGECRSCLENNCMRYIT-TCOPSSWSSVKNKIERFRKIYQF 181
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
87 KEKQALAEKQEVCEYNETMLSLWEECKPCLKITCMRVYSKICHSGSLVGRQLELLNRSPF 146

QY 182 LPFFHEDNEKDLPISEKLIK-----DAQLQMEDVFSQLTVDVNSLFNRSFNVFMQOQ 236
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
147 SIWNGERIDALLDRQROERFERDEERFGLMGVEDVIFODSTQLYCPAPFFR---- 202

QY 237 EFDQTFQSFISDPTLPEY--FFPAFSKEPMTKADLEOCWDIPNFQI-----FCNFS 288
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
203 -----TPFEGGFREAF-VPPVQRYRL-----VPPRRRLSRELHPFLQHP 240

QY 289 VSIYESVSSETITKMLKAIEDLPKQKADPH--GGILSKMLP-GODRGLCGELDONLSRCF 345
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
241 VGHGHLRFEMTQMLDG-----GHGANDHLLGGFESERNFSTDMVCRREIRNSAGCL 294

QY 346 KFEKCKQCOAHLSEDC-----PDVPAHLTELDEAIRLVNSNOQYGOILOMTRKHELEDTA 401
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
295 RMRDCEKCRREILAVDCSQTDPVQSQLREQFEDALRLAERTRRYDDLLSAFAQAEMLNTS 354

QY 402 YLVEKMRGQFGVWSELANOAPETEIFNSIQV---VPRIHGNTSKQDETMWTLISPLS 458
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
355 SLLDQLNQFGWLVRLGNLTQCTDGLQVTVTFVKTPNLEDP--SAPADTQVT-VQLFDS 411
```


DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CLUSTERIN (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE DAWLEY; TISSUE=THYMUS;
RA Park J.H., Park J.S., Ju S.K., Na S.Y., You K.H.;
RT "Determination of clusterin mRNA expression of apoptosis induced rat
thymocytes in vivo and in vitro."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR EMBL; AF314657; AAG31162.1; -.
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 2.
DR SMART; SM00035; Cla; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
KW Glycoprotein.
FT NON_TER 1 1
SQ SEQUENCE 218 AA; 24931 MW; ED2FC4425A510589 CRC64;

Query Match 7.8%; Score 204; DB 11; Length 218;
Best Local Similarity 28.3%; Pred. No. 1.3e-06;
Matches 47; Conservative 36; Mismatches 73; Indels 10; Gaps 3;
QY 330 DRGLGGLDQNLSCRFPHKCKQKQHLSEDC-----PDVPALHTELDEAIRLVNVSNOQ 385
Db 11 DTVCKEIRHSTGLCKMKGCKQCEILSYDCSTNNPAQANLQELNLSLQVAERLTQQ 110
QY 386 YGQIDQTKRKHLEDTAYLVKMRQFGWVSELAN--QAPETEIFNSIQVVPRTHEGNIS 443
Db 111 YNELLHSLSQKMLNTSLLEQLNDQFWSVQLANLTGDDOYLRSVTVT-----HSSDSE 166
QY 444 QDETMTDLSILPSSNTKLPLEESAESSNFYGVYVAKALQHF 489
Db 167 VPSRVTEYVVKLFDSPTTVVLPEEVSKDNPCKFMDTVAEKALQBYR 212

RESULT 9
O76329
ID O76329 PRELIMINARY; PRT; 1738 AA.
AC O76329
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INTERAPTIN.
GN ABPD.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98365468; PubMed=9700162;
RA Rivero F.J., Kuspa A., Brokamp R., Matzner M., Noegel A.A.;
RT "Interaptin, an actin-binding protein of the alpha-actinin superfamily
in Dictyostelium discoideum, is developmentally and CAMP-regulated and
RT associates with intracellular membrane compartments."
RL J. Cell Biol. 142:735-750(1998).
DR EMBL; AF057019; AAC34582.1; -.
DR HSSP; Q01082; 1BKR.
DR InterPro; IPR001589; Actinin_act_bind.
DR InterPro; IPR001715; Calponin_hom.

DR InterPro; IPR001990; Granin.
DR InterPro; IPR001451; Hexapep_transf.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00307; CH; 2.
DR SMART; SM00033; CH; 2.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE; PS00020; ACTININ_2; 1.
DR PROSITE; PS00021; CH; 2.
DR PROSITE; PS00422; GRANIN_1; UNKNOWN_1.
DR PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
SQ SEQUENCE 1738 AA; 204427 MW; 577A99D2EC79AF5C CRC64;

Query Match 6.2%; Score 164; DB 5; Length 1738;
Best Local Similarity 21.1%; Pred. No. 0.0087;
Matches 103; Conservative 87; Mismatches 171; Indels 128; Gaps 21;
QY 54 KDKTAISENL-----KSFSEVGEIDADEEVKKALTGKQKIMMER----- 94
Db 667 KDNQTINEQLKQSEKDEIKLSNQOQOQDEKINNLLLEKEKDCILIERINQQLLEN 726
QY 95 -----KEKHTNLMLSTLKKREEKQKALKLLNEVQEHLEEEER 132
Db 727 IDLSKYQQLLEFENFKLNSKEKE-----NQLNELQSKQDERENQLND--EKLEKEQ 779
QY 133 LCRESLADSWGECRSCLENNCMRYTTCQPSWSVKNKIERFFKYYQFLFPFHEDNEKD 192
Db 780 L--QSIEDFNQYK-----QQQLSSNSN-IDQQLQSTIIELSELKEQKELN 822
QY 193 LPISEKLIKDAQLTOMEDVFSOLT-----VDVNSLFNRSFNVRPMQOQEDDTFOSH 245
Db 823 ---DSKLLIEKQLQLOQEQFDFQLNEKNKXDQDQLELEKQ---LKQOQEQYDQLNETN 876
QY 246 FISDTLTPYFPAPFSKEPMTKADLE-----QCWDIPNFFQFCNFSVSIYESV 295
Db 877 QSIENQLNQNL---INKENLNEKEQELLKQLNQLNQIEKIQFDQOQFSKQNSINIELV 933
QY 296 SETITKMLKAIED---LPKQDKA---PDHGLISKMLPGQDRGLGGLDELONLSRCFKFHE 349
Db 934 NEKNEKLIQOQYDQLKQONRSNDKENDLIEK--ENOLKSIQNELNOLIEKNESDHK 991
QY 350 KQCKQAHLSQEDCPD---VPALHTELDE----AIRLVNVSNOQYGOIL-----Q 391
Db 992 EQQLKQOSIENDLIEKENIQQLQSQLNEQROQOSNOLSEKQDQQLNOLIEKNQFDQKEQ 1051
QY 392 MTRKHLDETAY----LVEKMRQGFQ--WVSELANQAPETEIFNSIQVVPRTHEGNISKQD 446
Db 1052 LKQOSIENDLFEKENIQQLQSQLNEQROQOSNOLSEKQDQQLNOLIEKNESDQK 1105
QY 447 ETMTDLSI 455
Db 1106 EQQLKQOSI 1114

RESULT 10
Q9W6V0
ID Q9W6V0 PRELIMINARY; PRT; 3616 AA.
AC Q9W6V0
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPERION PROTEIN, 419 KDA ISOFORM.
GN HYPERION.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RA Kemmer W.A., Schwarz U.;
RT "Characterization of Hyperion, a gene coding for an abundance of gene


```

SQ SEQUENCE 840 AA; 97587 MW; A875627BE3D4D716 CRC64;

Query Match
Best Local Similarity 5.7%; Score 149.5; DB 6; Length 840;
Matches 88; Conservative 89; Mismatches 154; Indels 143; Gaps 18;

QY 57 TAISENLKSFSEVGEIDAEVKKALGKQKIMMERKEKEHTNLTSLKKCRE-EKQE 115
DB 426 TEMQKNKSVSOY--LEMDKTLKKEEVKRLQLRKEQKVTAASADLLKREKETQE 483
QY 116 ALKLLNEVQE-----HLEEBELCRSLADSWGECRCLNENCMRIYTCQPSWSVKNKI 171
DB 484 FLSQEEFQKRDKNLEERQKL-----KSLR 509
QY 172 ERFRKIYQFLPFPHEDNEKDLPISEKLEKDAQLTOMEDVFSOLTVDVNSLNRSF--- 228
DB 510 EKLLTVKNLQFMSENERAKNIKQQ-----QINEVKNENKLLKHVARSEEQNVPK 562
QY 229 NVFQMOQEFDTQSHFISDITLTPYFFPAFS---KEPMTKADLE----- 272
DB 563 SETAQLKEQLAEVMSKSDITDKTMHSLLDSCPEESLNPAIDERSQLASKHSL 622
QY 273 -----QCWDIPNFQLFNCFNSIYSEVSEITMKAL-----EDLPKQDKAP 316
DB 623 ALMVGILLKQDITNSDAE--HFKESS--EKVSDIMLQRLKSLHLKKNLDRKLLKHKDRIT 678
QY 317 DHGGLISKMLPGQDRGL-----CGE----- 336
DB 679 TFRDLIAKAFQDHAKVTVDSCDEAKSIRDVPTFGAKLDKYHSLNEBELFLITYKLC 738
QY 337 -LDQNLRCFKFHEKQKQKQAHLSQEDCPDVPALHTLDEAIRLVNSNOQYQGIQWTR 395
DB 739 LLESKESHCHNLTENDKYRHGLSLIKVTVSEIEECADQRLAISHSQIAH-LEKRNK 797
QY 396 HLEDYALVEKMRGQFGWSELANOAPETELIEN---SIOVPRIHGNSKOD 446
DB 798 HLED---LIRKPR-----EKARK-PRSKSLENHFKSMTMPAVFENRNDLD 840

RESULT 13
Q28021 PRELIMINARY; PRT; 1388 AA.
AC 028021;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RHO-ASSOCIATED KINASE.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96208507; PubMed=8641286;
RA Matsui T., Amano M., Yamamoto T., Chihara K., Nakafuku M., Ito M.,
RA Nakano T., Okawa K., Iwamatsu A., Kaibuchi K.;
RT "Rho-associated kinase, a novel serine/threonine kinase, as a putative
RT target for small GTP binding protein Rho.";
RL EMBO J. 15:2208-2216(1996).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR HMBL; U36909; AAC48567.1; -.
DR HSSP; O63450; 1A06.
DR InterPro; IPR002219; DAG_PE-bind.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000961; Pkinase.C.
DR InterPro; IPR000861; REM_repeat.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00130; DAG_PE-bind; 1.
DR Pfam; PF02185; HRI; 1.
DR Pfam; PF00169; PH; 1.

DR Pfam; PR00069; pkinase; 1.
DR SMART; SM00109; CI; 1.
DR SMART; SM00074; HRI; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00133; S_TKC_X; 1.
DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE; PS00003; PH_DOMAIN; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1388 AA; 160799 MW; CA83CE7D3860465D CRC64;

Query Match
Best Local Similarity 5.7%; Score 149; DB 6; Length 1388;
Matches 95; Conservative 96; Mismatches 166; Indels 212; Gaps 20;

QY 54 KDKTAISENLKSFSEVGEIDAEVKKALGKQKIMMERKEKEHTNLTSLKKCRE-EKQE 104
DB 780 KQKDVNLNEDVRNLT-----LKIEQETQKRCITQNDLKMOTQOVNTLKMSEKQLKQENHLL 835
QY 105 TLK-----KCREKQKQKALKLLNEVQEHLEEE-----RLCR 135
DB 836 EMKMSLEKQNAELKQKQADQKQKELQQLAEQFSTLYKTQVRELKEBECEETKCLK 895
QY 136 E-----SLADSWGECRCLNENCMRIYTCQPSWSVKNKIERFRKIYQFLPFPHEDN 189
DB 896 ELQOKKQELQDE-----RDSLAQQLLEITLTKRADSEQLARSATGEYSDEKE 942
QY 190 E--KDLPISE-----KLEKDAQLTOMEDVFSOLTVDVNSLNFNRSPVFRQMOQEFQ 240
DB 943 KIMKELEIKEMARHKOELTEKDATIASLEETNRLTSDVANLANEKEELNKLKEAQSQ 1002
QY 241 TQSHFISDITLTPYFFPAFSKEPMTKADLEQCCWDIPNFQLFNCFNSIYSEVSEIT 300
DB 1003 LSR---LKDEISAAAKAQAQFEKOLLTERTLK-----TQAVN 1036
QY 301 KMLKATE-----DLPKQDKAPDHGGLISKMLPGQDRGLCGEL-----DQNLRCF 345
DB 1037 KLAETMNRKEPVKRGNDTVDVRKEK-----ENRKLHMLKSEREKLTKQMI 1082
QY 346 KFEKQKQKQAHLSQEDCPDVPALHTLDEAIRLVNSNOQYQGIQWTRKHL----- 398
DB 1083 KYQKELNEMQAQIAEE---SQRIELQMTLDSKSDIEQLRSQALHIGLDSSSISGS 1138
QY 399 -----DTRALVEKMRG-----QFGWSEL----- 417
DB 1139 PGDTEADGFPESRLEGWLSLPRNNTKFGWVKYVIVSVSSKKILFYDSEQDKESNPYM 1198
QY 418 -----ANQAPETELIFNSIOVVPRI-----HEGNIKQDETMMTDLSILPSSNFT 462
DB 1199 VLIDKLFHVRPVTQTDVIRADAKELPRIFQILYANEGESKEQE----- 1243
QY 463 LKIPLESAESSNFI---GYVAKALQHF 488
DB 1244 --FPVEPVGEKSNYICHKGHEFIPLYHF 1270

RESULT 14
Q18082 PRELIMINARY; PRT; 1091 AA.
AC Q18082;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 127.0 KDA PROTEIN.
GN C18C4.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:49:54 ; Search time 53.01 Seconds
(without alignments)
1556.661 Million cell updates/sec

Title: US-09-722-544A-4
Perfect score: 2529
Sequence: 1 MRTWDYNSGNMKPPLLVFI.....FIGYVAKALQHKEHFTW 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPTREMBL19.*
- 2: sp_archaea.*
- 3: sp_bacteria.*
- 4: sp_fungi.*
- 5: sp_human.*
- 6: sp_invertebrate.*
- 7: sp_mammal.*
- 8: sp_mhc.*
- 9: sp_organelle.*
- 10: sp_phage.*
- 11: sp_plant.*
- 12: sp_rodent.*
- 13: sp_virus.*
- 14: sp_vertebrate.*
- 15: sp_unclassified.*
- 16: sp_rvirus.*
- 17: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2464	97.4	466	4	Q15846
2	1836.5	72.6	465	6	Q95KN1
3	1230.5	48.7	338	6	Q9N1T8
4	423	16.7	449	6	Q29482
5	401.5	15.9	448	13	Q9YGP0
6	321	12.7	372	11	Q9JK98
7	266.5	10.5	295	11	Q35510
8	204	8.1	218	11	Q9ERD1
9	160	6.3	1738	5	Q76329
10	160	6.3	3616	13	Q9W6V0
11	151.5	6.0	1091	5	Q18082
12	149.5	5.9	840	6	Q95JRO
13	146	5.8	1156	16	Q66878
14	145	5.7	1388	6	Q28021
15	144	5.7	1305	10	Q9FJ35
16	144	5.7	3259	4	Q14789

17	143	5.7	1931	5	Q9NCF9	Q9ncf9 drosophila
18	143	5.7	1931	5	Q9VKH9	Q9vkh9 drosophila
19	142.5	5.6	333	6	Q95JY2	Q95jy2 macaca fasc
20	141	5.6	1388	4	Q95116	Q95116 homo sapien
21	141	5.6	1388	4	Q9UQN5	Q9uqn5 homo sapien
22	140.5	5.6	2166	16	O51465	O51465 borrelia bu
23	140	5.5	1379	11	O62868	O62868 rattus norv
24	139	5.5	1330	6	O97961	O97961 vulpes vulp
25	139	5.5	1530	4	O43241	O43241 homo sapien
26	139	5.5	1955	5	O61308	O61308 parascaris
27	139	5.5	5458	5	Q9U459	Q9u459 plasmodium
28	138.5	5.5	1391	11	Q922J3	Q922j3 mus musculu
29	137.5	5.4	720	4	O9H6Q7	O9h6q7 homo sapien
30	137	5.4	1084	16	O83423	O83423 treponema p
31	137	5.4	1300	4	O13999	O13999 homo sapien
32	136.5	5.4	709	5	O9GRG1	O9grg1 tetrahymena
33	136	5.4	1132	4	O75065	O75065 homo sapien
34	135.5	5.4	470	10	Q38843	Q38843 arabidopsis
35	135.5	5.4	684	3	Q07238	Q07238 pneumocysti
36	134.5	5.3	882	17	O96YR5	O96yr5 sulfolobus
37	134.5	5.3	1708	5	O9U0S6	O9u0s6 mytilus gal
38	134.5	5.3	1723	2	Q9JMX8	Q9jmx8 helicobacte
39	134.5	5.3	1958	5	O96062	O96062 dugesia jap
40	134	5.3	1356	4	Q14707	Q14707 homo sapien
41	133.5	5.3	1956	5	Q20641	Q20641 caenorhabdi
42	132	5.2	1033	4	O9UFE1	O9ufel homo sapien
43	132	5.2	1133	5	O21022	O21022 caenorhabdi
44	132	5.2	1270	4	O96JN2	O96jn2 homo sapien
45	132	5.2	1388	11	P70336	P70336 mus musculu

ALIGNMENTS

RESULT 1

Q15846	PRELIMINARY;	PRT;	466 AA.
ID Q15846;			
AC Q15846;			
DT 01-NOV-1996 (TREMELrel. 01, Created)			
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)			
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)			
DE HYPOTHETICAL 54.2 KDA PROTEIN PRECURSOR.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=NEURORETINA;			
RA Shimizu A., Nishida K., Kinoshita S., Inazawa J., Okubo K.,			
RA Matsubara K.;			
RT "Expression profile of active genes in human retina.";			
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.			
DR EMBL; D63813; BAA09882.1; -			
DR InterPro; IPR000753; Clusterin.			
DR Pfam; PF01093; Clusterin; 3.			
DR SMART; SM00035; CLa; 1.			
DR SMART; SM00030; CLb; 1.			
DR SIGNAL; Hypothetical protein.			
FT SIGNAL	1	20	POTENTIAL.
FT CHAIN	21	466	ROD PHOTORECEPTOR PROTEIN.
SQ SEQUENCE	466 AA;	54214 MW;	41A603D0FD6439C4 CRC64;

Query Match	97.4%	Score	2464;	DB	4;	Length	466;
Best Local Similarity	100.0%	Pred. No.	3.3e-161;				
Matches	466;	Conservative	0;	Mismatches	0;	Indels	0;
QY	12	MKPPLLVFVCLLWLKDSHCAPT	KDKT	TAISENLKSFSEVG	IDEADEVK	KALTG	KQMK 71
Db	1	MKPPLLVFVCLLWLKDSHCAPT	KDKT	TAISENLKSFSEVG	IDEADEVK	KALTG	KQMK 60
QY	72	IMMERKEHTNLMSTLKKCREEK	QEA	LKLLNEVOEHL	EEERLCRES	LADSWG	ECRSCL 131


```
Db 61 IMMERKEHTNLMSTLKKCKEKEQALKLLNEVQEHLEERLCRESLADSWGECRSL 120
Qy 132 ENNCNRIYTTCCPSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLTEDDAQLTQM 191
Db 121 ENNCNRIYTTCCPSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLTEDDAQLTQM 180
Qy 192 EDVFSQLTVDVNSLFRNRSFNFRMQOEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 251
Db 181 EDVFSQLTVDVNSLFRNRSFNFRMQOEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 240
Qy 252 DLEQCWDIPNFQFQFNFVSIVSVSTITKMLKAIEDLPKQKAPDHGGLISKMLPGQ 311
Db 241 DLEQCWDIPNFQFQFNFVSIVSVSTITKMLKAIEDLPKQKAPDHGGLISKMLPGQ 300
Qy 312 DRGLCGELDQNLSCRCFKHEKCKQCAHLSBDCPDVPALHTDELDAIRLVNVSNOOYGOI 371
Db 301 DRGLCGELDQNLSCRCFKHEKCKQCAHLSBDCPDVPALHTDELDAIRLVNVSNOOYGOI 360
Qy 372 LQMTKRKHLDTAYLVKMRGQFGWVSELANQAPETETIIFNSIQVVPRIHEGNISSKQDET 431
Db 361 LQMTKRKHLDTAYLVKMRGQFGWVSELANQAPETETIIFNSIQVVPRIHEGNISSKQDET 420
Qy 432 MTDSLILPSSNFTLKIPLSEASNSNFIYGVVAKALQHKHEKFTW 477
Db 421 MTDSLILPSSNFTLKIPLSEASNSNFIYGVVAKALQHKHEKFTW 466

RESULT 2
Q95KN1 PRELIMINARY; PRT; 465 AA.
AC Q95KN1
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE RETINAL CLUSTERIN-LIKE PROTEIN CLUL1B SPLICE VARIANT.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20156379; PubMed=10675623;
RA Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Aguirre G.D.;
RT "Molecular cloning, characterization and expression of a novel retinal
clusterin-like protein cDNA.";
RL Gene 243:151-160(2000).
DR EMBL: AF241221; AAK49030.1; -
SQ SEQUENCE 465 AA; 54402 MW; BBDEIAC512D5D33F CRC64;

Query Match 72.6%; Score 1836.5; DB 6; Length 465;
Best Local Similarity 73.4%; Pred. No. 3.5e-118;
Matches 342; Conservative 51; Mismatches 72; Indels 1; Gaps 1;

Qy 12 MKPPLLVTVCULLKDSHCAPTWKDKTAISENLKSFSEGEIDADEVKKAIGKQMK 71
Db 1 MKPSSLVFTVYLWLKDCHCAPTWKDKTMHGNLKGFEAGDIDVDEEVKKAIGKQMK 60
Qy 72 IMMERKEHTNLMSTLKKCKEKEQALKLLNEVQEHLEERLCRESLADSWGECRSL 131
Db 61 IMMERKEHTNLMSTLKKCKEKEQALKLLNEVQEHLEERLCRESLADSWGECRSL 120
Qy 132 ENNCNRIYTTCCPSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLTEDDAQLTQM 191
Db 121 ENSNCRHFTTCCPSWSSMKNVTVEQFERNIYQFLPPFEDNEKDLPVGEKFIEDDAQVAI 180
Qy 192 EDVFSQLTVDVNSLFRNRSFNFRMQOEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 251
Db 181 ENVFNQLTVDVRFNFRNLSUNVFKOMQOEFDQTFQSYFMSDTDLMPNLPALSKPEPRKA 240
Qy 252 DLEQCWDIPNFQFQFNFVSIVSVSTITKMLKAIEDLPKQKAPDHGGLISKMLPGQ 311
```

```
Db 241 DPVQSWDIPSFQFLFNFSLSIYHSITTTITKLTNAIEDLPKQDNDNSHGSLSKTLPVQ 300
Qy 312 DRGLCGELDQNLSCRCFKHEKCKQCAHLSBDCPDVPALHTDELDAIRLVNVSNOOYGOI 371
Db 301 HRGPGYGFQGNLSCFCQFHARCQCDYLWEDCPDELHTKVDLEALVNIHQQAQV 360
Qy 372 LQMTKRKHLDTAYLVKMRGQFGWVSELANQAPETETIIFNSIQVVPRIHEGNISSKQDET 431
Db 361 LQMTQHLEDITTYLMKMRREEFGVADLANQAPGAENIFDSTKMPNIEHGNFSKQDET 420
Qy 432 MTDSLILPSSNFTLKIPLSEASNSNFIYGVVAKALQHKHEKFTW 477
Db 421 I-DLSILSSNFTLKIPLSEASNSNFIYMLKAVQHFKHFKFTW 465

RESULT 3
Q9NLT8 PRELIMINARY; PRT; 338 AA.
AC Q9NLT8
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE RETINAL-SPECIFIC CLUSTERIN-LIKE PREPROTEIN.
GN CLUL1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20156379; PubMed=10675623;
RA Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Aguirre G.D.;
RT "Molecular cloning, characterization and expression of a novel retinal
clusterin-like protein cDNA.";
RL Gene 243:151-160(2000).
DR EMBL: AF147784; AAR36799.1; -
DR InterPro: IPR000753; Clusterin.
DR SMART: SM00035; CLA; 1.
DR SMART: SM00030; CLB; 1.
SQ SEQUENCE 338 AA; 39385 MW; B86F07877213FCC2 CRC64;

Query Match 48.7%; Score 1230.5; DB 6; Length 338;
Best Local Similarity 70.1%; Pred. No. 8.5e-77;
Matches 232; Conservative 41; Mismatches 57; Indels 1; Gaps 1;

Qy 147 SSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLTEDDAQLTOMEDVFSQLTVDVNSLF 206
Db 9 TSGTSTVEQFERNIYQFLPPFEDNEKDLPVGEKFIEDDAQVAQIENFNQLTVDVRLF 68
Qy 207 NRSFNVRMOMQOEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKADLCQCDWIDPINFOLF 266
Db 69 NRSNLVFKOMQOEFDQTFQSYFMSDTDLMPNLPALSKPEPRKADPVQSWDIPSFOLF 128
Qy 267 CNFSYIYSEVSETITKMLKAIEDLPKQKAPDHGGLISKMLPGODGLCGELDQNLSC 326
Db 129 NFNSLSIYHSITTTITKLTNAIEDLPKQDNDNSHGSLSKTLPVQHRPGYGFQGNLSEC 188
Qy 327 FKFEKCKQCAHLSBDCPDVPALHTDELDAIRLVNVSNOOYGOIQLMKRHLDTAYLV 386
Db 189 FQFHARCQCDYLWEDCPDELHTKVDLEALVNIHQQAQVQLQTHLEDITTYLM 248
Qy 387 EKMKGQFGWVSELANQAPETETIIFNSIQVVPRIHEGNISSKQDETMMTDLISLPSSNFTLK 446
Db 249 EKMKEFGVADLANQAPGAENIFDSTKMPNIEHGNFSKQDETI-DLSILSSNFTLK 307
Qy 447 IPLEESAESSNFIYGVVAKALQHKHEKFTW 477
Db 308 IPLEESAETSNSFIYMLKAVQHFKHFKFTW 338

RESULT 4
```

```

Q29482
ID Q29482          PRELIMINARY;      PRT;    449 AA.
AC Q29482;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CLUSTERIN PRECURSOR.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OLIVER; TISSUE=TESTIS;
RA Barber J.A., Farris J.A., Troedsson M.H.T., Crabo B.G., Foster D.;
RT "Nucleotide sequence of the complementary DNA encoding Equine Clusterin.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR EMBL; LA6797; AAA80313.1; -.
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; C1a; 1.
DR SMART; SM00030; C1b; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Signal.
FT SIGNAL                     1         21     POTENTIAL.
FT CHAIN                      22        449   CLUSTERIN.
SQ SEQUENCE 449 AA; 52154 MW; 705A60504515F7C0 CRC64;

Query Match              16.7%; Score 423; DB 6; Length 449;
Best Local Similarity 24.5%; Pred. No. 2.8e-21;
Matches 117; Conservative 100; Mismatches 203; Indels 58; Gaps

QY 16 LLYFIIVCLLMKSHCAPTWKDKATSENLKSFSEVGETADAEVKALTGIKQKMIMME 75
Db 4 LLLVLGULLFLENGQ---VLGDRAVDRELQENSTQGSNYINKEKNALKGVKKIKNLIE 60
QY 76 RKEKEHTNLMTLKCKREKQEKALLNEVOEHLEERLCRESLADSGEGRSCLENNC 135
Db 61 QTNEERKSLLGTLEEAKKKKEGALNTDKSEMKLKESQVCNETWTALWEECKPCLKTC 120
QY 136 MRYLT-TCQPSNSVNKRIERFRKIYQLFFPHEDNEKDLPISBKLIIEEDAQLTMEDV 194
Db 121 MKFYARVCRSGGLGVGHGLEEFNLQSSPFYFINGDR-----IDSLLENDRQQTHVLDV 174
QY 195 FSOLTVDVNSLFNRNFVRFMQOEFDQFQSHEFTSDTLTPPEYFPFAFSKPMTKADI- 253
Db 175 -----MQDSFDRASSI-----MDLFDQRFT-REPQTYIYSPPFS-SPHRRSSL 218
QY 254 -----EQCDIP-----NFFQLFCNFSVIYSVSSETITKMLKAIEDLPKQD--KAP 298
Db 219 FNPKSRFARNIMHPMYRHINFDMFPDPFDIMIHAQAAMNLHLHRLPDQLPMTFESGSD 278
QY 299 DHGGLLSKMLPGQDRGLCGELQDNLSRCFKFHEKCQKCOAHLSDC----PDVALHTEL 354
Db 279 NH-----DRTVCKEIRHNSTGCULKMKDCEKCEILLVDCSTNNPSQMQLROEL 327
QY 355 DEAIRLVNSNQOYGIOQLMTRKHLEDAYLVLEKMRGFGWGWSLANQAPETEIFNSIQ 414
Db 328 NNSLOLAEXFKTLKYDELLQSYQEKMLNTSSLKLQNLEQFSWVSQLANLTQGDDYYILOVT 387
QY 415 VVPRIEGHNTSKODETMMTDLSILSPSNFTLKIPLEBSAESNFTIGYVVAKALOHFKE 472
Db 388 TVSS-HNSDSEVPSGLTRVVVKLFOSPIYTVTPVEVSVRRNPKFMETVAERKALEYRQ 444

```

QY 432 MTDLSILPSSNFTLKIPLEESAESSNFGYVYVAKALQHFKH 473
Db 404 VT-VOLFSEPLSLVPGDISWDDPRFEIYVABQALQHYKQN 444

RESULT 6

Q9JK98 PRELIMINARY; PRT; 372 AA.
AC Q9JK98;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CLUSTERIN ISOFORM 2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=UTERUS;
RA You K.H., Jeon J.H.;
RT Identification of Truncated SGP-2 Lacking a Signal Peptide for
RT Secretion.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR EMBL; AF248058; AAF67185.1; -;
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; CLA; 1.
DR SMART; SM00030; CLB; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
KW Glycoprotein.
SQ SEQUENCE 372 AA; 43109 MW; A1047E48520FEE CRC64;

Query Match 12.7%; Score 321; DB 11; Length 372;
Best Local Similarity 23.6%; Pred. No. 2.2e-14;
Matches 93; Conservative 83; Mismatches 170; Indels 48; Gaps 12;

QY 95 KQALKLLNEVQBHEERLCRESLADSWGECRSCLENNCMRIYT-TCQPSWSSVYKNI 153
Db 4 REDALEDTRDSEMKLAFVEVCNETHMALWECKPCLKHTCKMFYARVCRSGSLVGQQL 63
QY 154 EREFRKIYQFLFPFHEDNEKDLPISKELTEDDAQLTQ----MEDVFSQLTVDVNSLFRNS 209
Db 64 EEFNLQSSPYFWMGDR-----IDSLLESDRQSQVLDAQMSFTRASGIIDTLFODR 117
QY 210 FNVFRMQQEFDTFQSHFISDFLTPEYFFP-----AFSKEPMTKADLEC-NDIPNF 262
Db 118 F--FAR-----ELHDPHYFS-----PFGPHKRPHEFLYPKSLRSLMSPSHYGPPSF 163
QY 263 FQLCFNFSYISVSSTITKMLKAIEDLPKQKAPDHGGLTSKMLPGQ--DRGLCGELDQ 321
Db 164 HNMFPFFEMTHQAQAMVQLHSPAFQPPD-----FLREGEDDRTVCKEIRR 213
QY 322 NLSRCFKHEKCKQKQAHLSDC-----PDVPALHTELDEAIRLVNVSNOOYGQILQMTKR 377
Db 214 NSTGCLMKQCKECCOEIILVDCSTNNPAQANLRQELNDSLQVAERLTQYKELLOSFOFS 273
QY 378 HLEDATYLVKMRGQFQVSWSELANQAPETIIFNSIQVVPRIHEGNTSKQDETMTDLISI 437
Db 274 KMLNTSLLLEQLNDQFNWSQLANLTQGEDKYLYRSTV--TTSSDSEVPSTVEVVVKL 332
QY 438 LPSGNFTLKIPLEESAESSNFGYVYVAKALQHFK 471
Db 333 FDSPTITVLPVEYSKDNPKFMDTVAEKALQEYR 366

RESULT 7

Q35510 PRELIMINARY; PRT; 295 AA.
ID Q35510;
AC Q35510;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CLUSTERIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR; TISSUE=SEMINAL VESICLE;
RX MEDLINE=98163873; PubMed=9503143;
RA Izawa M.;
RT Identification of a transcript predicting an alternative form of
RT sulfated glycoprotein-2 (clusterin) in rat tissues.;
RL Biochem. Mol. Biol. Int. 44:9-19(1998).
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR EMBL; D11339; BAA21775.1; -;
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; CLA; 1.
DR SMART; SM00030; CLB; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
KW Glycoprotein.
SQ SEQUENCE 295 AA; 34115 MW; C1E07FA745DD5470 CRC64;

Query Match 10.5%; Score 266.5; DB 11; Length 295;
Best Local Similarity 25.2%; Pred. No. 9.2e-11;
Matches 80; Conservative 59; Mismatches 131; Indels 47; Gaps 11;

QY 115 LCRESLADSWGECRSCLENNCMRIYT-TCQPSWSSVYKNIERFKIYQFLFPFHEDNEK 173
Db 9 VCNETHMALWECKPCLKHTCKMFYARVCRSGSLVGRLQLEEFNLQSSPYFWMGDR-- 66
QY 174 DLPISKELTEDDAQLTQ----MEDVFSQLTVDVNSLFRNSFVRMQQEFDTFQSHFI 229
Db 67 ----IDSLLESDRQSQVLDAQMSFTRASGIIDTLFODR--FTHEPQDI-----HHF- 114
QY 230 SDTDLTEPYFFP-----AFSKEPMTKADLECQNDIP--NFQLCNFSVSIYSVSETIT 282
Db 115 -----SPMGFPHKRPHEFLYPKSLRSLMPLSHYGLPSLHFMFPFPFDMTHQAQAMDV 168
QY 283 KMLKAIEDLPKQKAPDHGGLTSKMLPGQ--DRGLCGELDQNLRCFKHEKCKQKQAHLS 341
Db 169 QJHSPALQFPDQVDFLKE-----GEDDTVCKEIRHNSTGCLMKMGQCKQCEILS 218
QY 342 EDC-----PDVPALHTELDEAIRLVNVSNOOYGQILQMTKRKHELTAYLVKMRGQFGWS 397
Db 219 VDCSTNNPAQANLRQELNDSLQVAERLTQYVNEILLSQSKMLNTSLLEQLNDQFSWVS 278
QY 398 ELANQAPETIIFNSIQ 414
Db 279 QLANTLQGGDDQYFGSPQ 295

RESULT 8

Q9ERD1 PRELIMINARY; PRT; 218 AA.
ID Q9ERD1
AC Q9ERD1;

RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.,
RT Isolation of novel full-length cDNA clones from macaque testis cDNA
libraries.,
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB070119; BAB63064.1; -
KW Hypothetical protein.
SQ SEQUENCE 840 AA; 97587 MW; A875627BE3D4D716 CRC64;

Query Match 5.9%; Score 149.5; DB 6; Length 840;
Best Local Similarity 18.6%; Pred. No. 0.034;
Matches 88; Conservative 89; Mismatches 154; Indels 143; Gaps 18;

QY 39 TAISENLKSFSEGEIDAEVKKALGIGIKIMMERKEHEHTNLMSTLKKCRE-EKQE 97
DB 426 TEMQKNKSVSQY--LEMDKTSKKEEVKRLQOLRKEQEVKTASALDILKKKEKTEQOE 483
QY 98 ALKLLNEVOE---HLEEEERLCRESLADSWGCRSCLENNCMRIYTTCCPSWSSVKNKI 153
DB 484 FLSLQEEFKRDKANLEERQKL-----KSLR 509
QY 154 ERFRKIYQFLFPFHEDNEKDLPISEKLEEDAQLTQMEDVFSQLVNLSFNRSF---- 210
DB 510 EKLTVQKNLQFMSENERAKNIKLOQ-----QINEVKNENKDKQHVARSEQNVYVK 562
QY 211 NVFRMQQEFDDTFQSHFISDTLTETPYFFPAFS---KEPMTKADLE----- 254
DB 563 SETAQLKEQLEEVYMKSDITKDTKMTNSLLDCSPCEESLNPAIDERSQLASKMHSLL 622
QY 255 -----QCWDIPNFFQCFNFSVSVSEYSETITKMLKAI-----EDLPKQDKAP 298
DB 623 ALAVGLLKCDITNSDAE--HFKESS--EKVSDIMQLRKLKSLKLNKDKELLKHDRIT 678
QY 299 DHGGLISKMLPGQDRL-----CGE----- 318
DB 679 TFRDLIAKEAFQDHAIKVTDCDSDEAKSIRDVPTFLGAKLDKYHSLNEELDLITKLGC 738
QY 319 -LDQNLRSRCPKFEKQKCOAHLSDCPDVPAHLHTLDEALRVNVSNOQYQILOWTRK 377
DB 739 LLESKESHCHNRLLLENDKYORHGLSLKKVTSYEEIECADQRLAISHSQIAH-LEKRNK 797
QY 378 HLEDYATLVKMRGQFGWSELANQAPETIIFN-----SIQVVPRIHEGNTSKOD 428
DB 798 HLED---LIRKPR-----EKARK-PRSKSLENHPKSMTPMVAPEKFNRLD 840

RESULT 13
O66878 PRELIMINARY; PRT; 1156 AA.
AC O66878;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CHROMOSOME ASSEMBLY PROTEIN HOMOLOG.
GN XCP OR AQ_629.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus";
RL Nature 392:353-358(1998).
DR EMBL; AE000699; AAC06839.1; -
DR InterPro; IPR003439; ABC_Transportr.
DR InterPro; IPR001687; ATP_GTP-A.
DR InterPro; IPR003716; RNA_pol_omega.

DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF02463; SMC_N; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1156 AA; 135563 MW; B12DB30F70C0CE49 CRC64;

Query Match 5.8%; Score 146; DB 16; Length 1156;
Best Local Similarity 20.0%; Pred. No. 0.086;
Matches 89; Conservative 89; Mismatches 162; Indels 104; Gaps 17;

QY 45 LKSFSEVGEID---ADEVKKALGIGIKIMME-----RKEHEHTNLMSTLKK 90
DB 163 IEISGIGYEYERKKALEELAEVELKKEIDILIEESNQLKRLKEKEKLEKELQOR 222
QY 91 CREEKQEAALKLLNEVQEHLEEEERLCRE--SLADSWGCRSCLENNCMRIYTTCCPSWSS 148
DB 223 IKRE-TEAKILLKEKEKLLKERITLNELSSLSRESLEDITFOIQENKEL----- 271
QY 149 VNKIERFRKIYQFLFPFHEDNEK---DLPISEKLI-EEDAQLTQMEDVFSQLVNVS 204
DB 272 --NERERLLKEYNEKIMPFKEKVGKFTABEIAERSIKEKEKELKESENVRNKLLEINN 329
QY 205 LFNRSFNVRQ--MOQEFDDTFQSHFISDTLTETPY-FFPAFSKEPMTKADLEQCWDIP 260
DB 330 LLSKRENLEREVGTLQLELEK-----LKEYSKLSKEVERKLELEEEE----- 373
QY 261 NFFQLCFNF-SVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQDRLGCGEL 319
DB 374 --ERLKITFDEVKLEKEKEKLEKLNLSL-NKEKQELEIQRANLNKI-----ERI 421
QY 320 DONLSRCFFHEKQKCOAHLSDCPDVPAHLHTLDEALRVNVSNOQYQILOWTRKHL 379
DB 422 KEDINKLISEREKEIKKEIKKEIKKAIKKKEEELRNLTQELNIYEKRLSEVRKKL 481
QY 380 E-----DTAVLVEKMRGQFGWVSELAN-QAPE----- 405
DB 482 EEVLKEKGAIEREVRFSFSDVFKDIKGVGVSVELIRVKNPEHITAIEVAGGRLKFI 541
QY 406 ----TEIIFNSIQVVPRIHEGNIS 425
DB 542 VVEDEEVAKECIQAKRMNMGFS 565

RESULT 14
Q28021 PRELIMINARY; PRT; 1388 AA.
AC Q28021;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RHO-ASSOCIATED KINASE.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96208507; PubMed=8641286;
RA Matsui T., Amano M., Yamamoto T., Chihara K., Nakafuku M., Ito M.,
RA Nakano T., Okawa K., Iwamatsu A., Kaibuchi K.;
RT "Rho-associated kinase, a novel serine/threonine kinase, as a putative
target for small GTP binding protein Rho.";
RL EMBO J. 15:2208-2216(1996).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; U36909; AAC48567.1; -
DR HSSP; Q63450; 1A06
DR InterPro; IPR002219; DAG_PE-bind.
DR InterPro; IPR00719; Euk_pkinase.

DR InterPro: IPR001849; PH.
DR InterPro: IPR000961; PKinase_C.
DR InterPro: IPR000861; REM_repeat.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR Pfam: PF00130; DAG_PE_bind; 1.
DR Pfam: PF02185; HR1; 1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00069; PKinase; 1.
DR SMART: SM00109; C1; 1.
DR SMART: SM00074; HR1; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00220; S_TKC; 1.
DR SMART: SM00133; S_TKC_X; 1.
DR PROSITE: PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR KW ATE-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1388 AA; 160799 MW; CAB3CE7D3860465D CRC64;

Query Match 5.7%; Score 145; DB 6; Length 1388;
Best Local Similarity 16.5%; Pred. No. 0.12; 166;
Matches 94; Conservative 97; Mismatches 166; Indels 212; Gaps 20;

QY 36 KDKTAISENLKSFSEVGIDEAEVVKALTGKIKOMKI-----MMERKEKEHTNLM 86
DB 760 KOKVLNEDVRNLT-----LKIEQETQRCLTQNDLMKMQVNTLKMSEKQLKQENHLL 835
QY 87 TLK-----KCREKEQALKLNEVQEHLEEE-----RLCR 117
DB 836 EKMSLEKQNALRKERDQACQMKELQDQAEQYFTLYKTVRELKECEKTKLCK 895
QY 118 E-----SLADSWGECRSCLENNCMRIYTTQCPSSVSKNKKIERFFRKIYFLFPFHEDN 171
DB 896 ELQKKQELQDE-----RDSLAQLETLTKADSEQLASIAEQSDLEKE 942
QY 172 E--KDLPISE-----KLIEDAQLQTMEDVFSQTLVDVNSLFRNFVFMQOQEPDQ 222
DB 943 KIMKELEIKEMMARHKOELTEKDATIASLEETNRTLSDVANLANEKEELNKLKEAQEQ 1002
QY 223 TFQSHFISDTLTETFPYPAFSKPMTKADLEQCWDIPNFFQCNFVSIVYESVSETIT 282
DB 1003 LSR---LKDEISAAAKAQEKOLLERTLK-----TQAVN 1036
QY 283 KMLKAIE-----DLPQKAPDHGGLISKMLPGDGRGLGEL-----DONLSRCF 327
DB 1037 KLAETMNRKEPVKRGNDTDVRRKEK-----ENRKLHMLKSEREKLITQOMI 1082
QY 328 KFHEKCKCOAHLSEDCPDVPALHTELDEAIRLVNSNQYQGLQWTRKHL-----380
DB 1083 KYQELNEMQAIQAE-----SQRIEQLMTLDSKSDIEQLRSQALHIGLSDSSIGSG 1138
QY 381 -----DTAYLVEKMRG-----QFGWVSEL-----399
DB 1139 PGDTEADGGPPEKSLGLWLSLPVRNNTKKGWKKYIVVSSKTLIFVDSQDKEQSNPYM 1198
QY 400 -----ANQAPETELIIFNSIQVVPRI-----HEGNTSKQDETMTDLSTLPSSNFT 444
DB 1199 VLDIDKLFLHVRPVQTQDVYRADAKEIRIFQILYANEGESKKEQ-----1243
QY 445 LKIPLEESAESSNFI---GYVAKALQHF 470
DB 1244 --FPVEPVGEKSNVICHGHEFTPLHF 1270

RESULT 15
Q9FJ35
ID Q9FJ35 PRELIMINARY; PRT; 1305 AA.
AC Q9FJ35;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MYOSIN HEAVY CHAIN-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=99087489; PubMed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned PI and TAC clones.";
RL DNA Res. 5:297-308(1998).
DR EMBL; AB016871; BAB10654.1; .
SQ SEQUENCE 1305 AA; 149961 MW; F37284B8E7BBF217 CRC64;

Query Match 5.7%; Score 144; DB 10; Length 1305;
Best Local Similarity 18.8%; Pred. No. 0.14;
Matches 94; Conservative 82; Mismatches 176; Indels 148; Gaps 18;

QY 26 LKDSHCAPTWKDKTAISENLKSFSEVGIDEAEVVKALTGKIKOMKIMMERKEKEHTNLM 85
DB 127 LKESHV---KERELFS---LRDIHEIHQRDSS-----TRASELAQLLESSKQOVSLS 174
QY 86 STLKKKREKQEAULKLENEVQEHLEEBERLCRESLA-----DSWGECRSCLENNCMRIY 139
DB 175 ASLKAABEENKAISKNVETMKNLEQTQNTIQELMAELGKLDKSHREKESLS-----228
QY 140 TTCOPSSVSKNKKIERFFRKIYQFLFPFHEDNEKDLPISEKLEED-----AQLTQM 191
DB 229 -----LVEVETHORDSSIHVKELEEQYESSKKLVAELNQT 264
QY 192 -----EDVFSQTLVDVNSLFRNFVFMQOQEPDQTFQSHFISDTLTETFPYPAFS 244
DB 265 LNNAEEKKVLKSQIAELSNELKEAQTQIELYSESQLKESHVKDRDL-----FS 316
QY 245 -----KEPMTK--ADLEQCWDIPNFFQCNFVSIVYESVSETITKMLKAIEDLPKQ 294
DB 317 LRDIETHQRESSTRVSELEA-----QL-----ESSQRLSDLTVDLKDAEEE 359
QY 295 DKAPDHGGLISKMLPGDGRGLGELDONLSRCFKFHEKCKCOAHLSEDCPDVPALHTEL 354
DB 360 NKA-----ISKNLKIMDK-----LEQAQNTIKELMDELGELKDRHKEK 398
QY 355 D-BAIRLVNSNQYQGLQWTRKHLLEDYALVKEKMRGQFGWVSELANOAPET-----406
DB 399 ESELSSLVKSADQOQVADMKOSL-----DNAEEKKMLSQ--RILDISNETQEAQKTIQEH 451
QY 407 -----EIIFNSIQVVPRIHEGNTSKQDETMTDLSTLPSSNFTKIPLEES 452
DB 452 MSSEQLKESHGKVERELTGLRDIETHQRESSTRLESETQLKLEQRVVDLSASLNAA 511
QY 453 AESSNFIQYVAKALQHFKE 472
DB 512 EEKKSLSLSMILEITDELQK 531

Search completed: July 2, 2002, 11:49:57
Job time: 391 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:51:19 ; Search time 56.51 Seconds
(without alignments)
937.572 Million cell updates/sec

Title: US-09-722-544A-4MOD
Perfect score: 2529
Sequence: 1 MRTWDSYSGNKKPLLVFI.....FIGYVAKALQHFKEHFKTW 477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2485	98.3	477	20 AAY30786	Protein encoded by
2	2460	97.3	466	20 AAY30793	Immature human HKN
3	2437	96.4	495	20 AAY30785	Protein encoded by
4	2346	92.8	446	20 AAY30792	Mature secreted hum
5	1753.5	69.3	465	20 AAY30791	Bovine HKNG1 ortho
6	1635	64.7	466	20 AAY30787	Protein encoded by
7	1554	61.4	450	20 AAY30788	Guinea pig HKNG1 o
8	1277	50.5	374	20 AAY30789	Guinea pig HKNG1 o
9	1272.5	50.3	373	20 AAY30790	Guinea pig HKNG1 o
10	780.5	30.9	521	20 AAY30794	Amino acid sequenc
11	434	17.2	449	22 AAU28048	Novel human secret

12	432	17.1	448	12	AAE11704	Cytolysis Inhibito
13	405.5	16.0	416	22	ABB50285	Apolipoprotein J o
14	400.5	15.8	446	21	AAE03441	Porcine clusterin
15	320	12.7	363	22	AAE03764	Human gene 1 encod
16	173.5	6.9	148	22	AAO00607	Human polypeptide
17	160.5	6.3	116	21	AAE03745	Human secreted pro
18	159.5	6.3	139	22	AAE03783	Human gene 1 encod
19	159.5	6.3	139	22	AAE01703	Human gene 4 encod
20	152.5	6.0	247	22	AAE03787	Human gene 1 encod
21	149	5.9	1372	19	AAE56473	Protein with Rho p
22	145	5.7	944	21	AAE67600	Human adipose tiss
23	145	5.7	1388	19	AAE56475	Protein with Rho p
24	143	5.7	1931	22	ABE61012	Drosophila melanog
25	141.5	5.6	3433	18	AAE22017	Utrrophin. Homo sa
26	140	5.5	934	22	AAU01768	Human secreted pro
27	140	5.5	2517	21	AAE71159	Human phosphodiester
28	139.5	5.5	2688	22	AAE40883	Human polypeptide
29	139	5.5	2633	22	ABG06505	Novel human diagno
30	138.5	5.5	2663	22	AAE39097	Human polypeptide
31	138	5.5	5373	22	AAU14603	Novel bone marrow
32	138	5.5	5447	22	AAU14697	Novel bone marrow
33	135.5	5.4	1427	12	AAE10534	Human 160kD mediat
34	135	5.3	1374	22	AAE69070	Human male enhance
35	133	5.3	990	22	AAE78520	Human protein SEQ
36	131.5	5.2	1411	17	AAE02258	Nucleolar/endosoma
37	130.5	5.2	2482	16	AAE72826	Human mitosis. Ho
38	130.5	5.2	2482	19	AAE23996	Human mitosis amin
39	130	5.1	1788	22	AAE40467	Human polypeptide
40	130	5.1	3248	17	AAE9795	Kinetochore protei
41	129	5.1	931	22	AAE79504	Human protein SEQ
42	128.5	5.1	1392	20	AAE06999	Restin protein seq
43	127.5	5.0	721	21	AAE21227	Protein encoded by
44	127.5	5.0	2053	22	AAU03501	Human protein kina
45	127	5.0	1047	18	AAE01535	Cellular homologue

ALIGNMENTS

RESULT 1
AAY30786
ID AAY30786 standard; Protein; 477 AA.

AC AAY30786;

DT 23-NOV-1999 (first entry)

DE Protein encoded by a human HKNG1 splice variant HKNG1-V1.

KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW schizophrenia; splice variant.

OS Homo sapiens.

PN WO9947535-A1.

PD 23-SEP-1999.

PF 16-MAR-1999; 99WO-US05606.

PR 16-MAR-1998; 98US-0078044.

PR 05-JUN-1998; 98US-0088312.

PR 28-OCT-1998; 98US-0106056.

XX 22-JAN-1999; 99US-0236134.

PA (MILL-) MILLENNIUM PHARM INC.

XX (REGC) UNIV CALIFORNIA.

PI Chen H, Freimer NB;

XX WPI; 1999-562047/47.

DR N-PSDB; AA210751.

XX New HKNG1 polynucleotides useful in diagnosis and treatment of
PT neuropsychiatric disorders, e.g. bipolar affective disorders and
PT schizophrēnia
XX
XX Claim 1; Fig 2A-B; 205pp; English.
XX
XX The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)
XX splice variant. HKNG1 is a gene associated with bipolar affective
XX disorder (BAD). HKNG1 polynucleotides are useful to identify compounds
XX modulating HKNG1 gene expression or HKNG1 polypeptide expression/
XX activity. Compounds inhibiting or enhancing HKNG1 gene expression or
XX activity in individuals can then be administered therapeutically to
XX treat HKNG1-mediated disorders, especially neuropsychiatric disorders
XX e.g. BAD, schizophrēnia, or HKNG1-mediated myopia disorders, such as
XX early-onset autosomal dominant myopia. The polynucleotides can be used
XX in gene therapy techniques to treat such disorders. They are also useful
XX in diagnosis to identify individuals having, or at risk of developing,
XX HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such
XX mutations especially result in the production of a protein with a
XX different sequence to the human full-length HKNG1 polypeptide or
XX splice variant sequences, especially the substitution of a lysine for
XX a glutamic acid at residue 202 or 184. The polynucleotides are also
XX useful in gene mapping, to produce probes or primers to identify
XX similar sequences (e.g. mutants or sequences from different species)
XX and to produce transgenic animals.
XX
SQ Sequence 477 AA;

Query Match 98.3%; Score 2485; DB 20; Length 477;
Best Local Similarity 99.2%; Pred. No. 4.8e-201;
Matches 472; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MRTDWSNGNKPPLLYFIVCLLWKDSCAPTWKDKTAISENLKSFSEVGEIDADEEV 60
DB 1 mrtdwsngnkppllyfivcllwkshc aptwkdk taise nksfsevg eidaadeev 60
QY 61 KKAALGKQKIMMERKEHEHTNLMTLKKCREEKOALKLNEVOEHLEEBEERLCRESL 120
DB 61 kkaalgkqkimmerkehtnlmtlkkcreekqealkllnevehleeeerlcresl 120
QY 121 ADSWGECSRLENMCMRYTTTCOPSSVKNKIERFFRKIYQFLPFPHEDNEKDLPISEK 180
DB 121 adswgecsr lenmcmrytttcopssvknkie rffrk i yqfl pfphe dnekd lpi sek 180
QY 181 LIEKDAQLTQMEDVFSQLTVDVNSLFNSFNFRMQQEEFDTQSFHSIDTDLTEPYFF 240
DB 181 liekdaqltqmedv fsq ltv dvn s lfn sfn frmq qe fdt qsf hsi dtd lte pyff 240
QY 241 PAFSKEPMTKADLEQWDIPNFQFCNFSVSIYSEVSETITKMLKATIEDLPKQKAPDH 300
DB 241 pafskepmtkadleqwdipnfqfcnfsvsiy sesvsetitkmlka tiedlpkqk apdh 300
QY 301 GGLSKMLPGQDRGLCGELDONLSRCFRFHEKQCQAHLSBDCQDPVPALHTELDEAIRL 360
DB 301 gglskmlpgqdr glc gel don lsr cfr fhe kqc qah lsb dcq dpv pal htel de air l 360
QY 361 VNVSNQYQGIQLQMTKRKLEDAYLVEKMRGQFGWSELANOAPETETIFNSIQVVPRIH 420
DB 361 vnvsnqyqgiqlqmtkrkledaylvekmrgqfgwselan oapet etifn siqv vpr ih 420
QY 421 EGNISKQDETMTDLSILPSSNFTLIKIPLEESAESNSFTIGYVVAALQHFKEHFT 476
DB 421 egniskqdetmtdlsilpssnftlikipleesaesnsftigyvvaalqhfkehft 476

RESULT 2
RAY30793
ID AAY30793 standard; Protein; 466 AA.
XX
AC AAY30793;
XX

DT 23-NOV-1999 (first entry)
XX
DE Immature human HKNG1 protein form 2.
XX
KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW schizophrēnia; splice variant.
XX
OS Homo sapiens.
XX WO9947535-A1.
XX
XX 23-SEP-1999.
XX
XX 16-MAR-1999; 99WO-US05606.
XX
XX 16-MAR-1998; 98US-0078044.
XX 05-JUN-1998; 98US-0088312.
XX 28-OCT-1998; 98US-0106056.
XX 22-JAN-1999; 99US-0236134.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX (REGC) UNIV CALIFORNIA.
XX
XX Chen H, Freimer NB;
XX
XX WPI; 1999-562047/47.
XX
XX New HKNG1 polynucleotides useful in diagnosis and treatment of
XX neuropsychiatric disorders, e.g. bipolar affective disorders and
XX schizophrēnia
XX
XX Claim 22; Fig 17; 205pp; English.
XX
XX The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1
XX is a gene associated with bipolar affective disorder (BAD). HKNG1
XX polynucleotides are useful to identify compounds modulating HKNG1 gene
XX expression or HKNG1 polypeptide expression/activity. Compounds inhibiting
XX or enhancing HKNG1 gene expression or activity in individuals can then
XX be administered therapeutically to treat HKNG1-mediated disorders,
XX especially neuropsychiatric disorders e.g. BAD, schizophrēnia, or
XX HKNG1-mediated myopia disorders, such as early-onset autosomal
XX dominant myopia. The polynucleotides can be used in gene therapy
XX techniques to treat such disorders. They are also useful in diagnosis
XX to identify individuals having, or at risk of developing, HKNG1-mediated
XX disorders due to mutations in the HKNG1 gene. Such mutations especially
XX result in the production of a protein with a different sequence to
XX the human full-length HKNG1 polypeptide or splice variant sequences,
XX especially the substitution of a lysine for a glutamic acid at residue
XX 202 or 184. The polynucleotides are also useful in gene mapping, to
XX produce probes or primers to identify similar sequences (e.g. mutants
XX or sequences from different species) and to produce transgenic
XX animals.
XX
SQ Sequence 466 AA;

Query Match 97.3%; Score 2460; DB 20; Length 466;
Best Local Similarity 99.8%; Pred. No. 5.9e-199;
Matches 465; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 MKPPLLVTVCCLLWKDSCAPTWKDKTAISENLKSFSEVGEIDADEEVKALTGIKQMK 71
DB 1 mkppllvtvccllwkshc aptwkdk taise nksfsevg eidaadeevkaltgikmqk 60
QY 72 IMMERKEHTNLMTLKKCREEKOALKLNEVOEHLEEBEERLCRESLADSWGCSRCL 131
DB 61 immerkehtnlmtlkkcreekqealkllnevehleeeerlcresl adswgcsrcl 120
QY 132 ENNCMRIYTTTCOPSSVKNKIERFFRKIYQFLPFPHEDNEKDLPISEKDLQLTQM 191
DB 121 enncmriytttcopssvknkie rffrk i yqfl pfphe dnekd lpi sek dlt qm 180

Qy	192	EDVFSQLTVDVNSLFNSFNVPQMQOEFDTQTSUHFISDITLUTEPYFFPFSKEPMTKA	251
Db	181	edvfsqltvdvnslnfsnfvrmqmgefqdtqgshfisdtdltepyffpfskepmtka	240
Qy	252	DLEQCWDIPNFFQFCNFNFVSIVESVETITKMLKAIEDLPKODKAPDHGGGLSKMLPQ	311
Db	241	dleqcwdipnffqfcnfsvsivesvettikmlkaiedlpkqdkapdhgggliskmlpq	300
Qy	312	DRGLCGELDONLSRCFKFHEKCKQCAHLSDECDPVDPAULHTELDEAIRLVNVSNOQYGOI	371
Db	301	drglcgeidgnlsrctfkfhekcckqahlsedcdpvpahlteldeairlvnvsngqygi	360
Qy	372	LQMTRKHELDATYLVKMKRGQFGWSELANOAPETEIFNSIQVVPRIHEGNISKODETM	431
Db	361	lqmrkthledatyvekmrgqfgwselanqapeteiifnsiqvvprihegniskqdeum	420
Qy	432	MTDLSILPSSNFTIKTLPDESASSNFYGVAKALQHFKEFKTW	477
Db	421	mtdslpssnftiklplessaessnfgygvakalqhfkefkwt	466

RESULT	3
AAAY30785	
ID	AAAY30785 standard; Protein; 495 AA.
XX	
XX	AAAY30785;
XX	
DT	23-NOV-1999 (first entry)
XX	
XX	Protein encoded by human HKNG1 cDNA.
DE	
XX	
XX	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW	schizophrenia; splice variant
KW	

disorders due to mutations in the HKNG1 gene. Such mutations especially result in the production of a protein with a different sequence to the human full-length HKNG1 polypeptide or splice variant sequences, especially the substitution of a lysine for a glutamic acid at residue 202 or 184. The polynucleotides are also useful in gene mapping, to produce probes or primers to identify similar sequences (e.g. mutants or sequences from different species) and to produce transgenic animals.

Sequence 495 AA;

Query Match	96.43;	Score 2437;	DB 20;	Length 495;
Best Local Similarity	98.7%;	Pred. No. 5.6e-137;		
Matches 464;	Conservative 2;	Mismatches 4;	Indels 0;	Gaps 0;
Qy	7	SNSGNMKPPLLVFFVTVCLLWLKDSHCAPTWKDKTATSENLSKFSFSEVGETDADEVKKALTG	66	
Db	25	nsngnmkppllvfvclvklwdkshcaptwkdkttaisenlksfsevgeldadeevkaltg	84	
Qy	67	IKQWKIMMERKEKEHTNLMSTLTKCKREBKQALKLLNVEQHLBEEERLCRESLADSWGE	126	
Db	85	ikgmkimmerkekehtnlmstlkkcreekqaalkllnevqehleeeerlcrseldadshge	144	
Qy	127	CRSLENNCMRIYTCOPSSWSNVKNKTERFRKLYQFLPFRHEDNEKDLPISEKLTIEKDA	186	
Db	145	crslenncmriytcopsswvknkterfrkiyqflfpfhednekdlpisekllieeda	204	
Qy	187	QLTOMEDVFSQLTVDVNSLFRNRSNVRFMQOEFDQTFQSHFISDTLTEPYFFPAFSKE	246	
Db	205	qltkedvfgsltvdvnslnfrsnfvrkqgepdqtfqshfisdtdltepyffpafske	264	
Qy	247	PMTKADLEQCCWDIENFQLCNFSVSIVSEYSEITKMLKAIEDLPKODKAPDHGGGLSK	306	
Db	265	pmtkadleqcdwipofqglfcnfsvsiyeseitkmlkaiedlpqdkapdhggglisk	324	
Qy	307	MLPQDGRGLGELDQNLSPCKPFHEKCKOKCOAHLSDECPVOPVALHTELDEAIRLVNVSQ	366	
Db	325	mllpgqdrglgclgdqnlsrcfkhckqckqahlsedcpvpalhteleidairlvnvsq	384	
Qy	367	QYGQILQMTKRKHLEDATYLVKMKRGQFGWSELANOAPETEIFNSIQVVPRIHEGNLSK	426	
Db	385	qygqilqmtkrkhledatylvekmrgqfgwselanqapeteiifnsiqvvprihegnlsk	444	
Qy	427	QDEFTMTDLSILPSSNTLTKIPLBESAESSNFICYVVAQALQHKRKEHFT	476	
Db	445	qdeumtcldsilpssnftlkiplbeessnfiqyvvaqalqhrkehft	494	

RESULT	4
RAY30792	
ID	RAY30792 standard; Protein; 446 AA.
XX	
AC	AAV30792;
XX	
XX	
DT	23-NOV-1999 (first entry)
XX	
DE	Mature secreted human HKNG1 protein sequence.

Db 1 mkppllvfviylqlrdccqcapgkgrtsiredpksfsgageidvdeevkkaaligmkqm 60
 QY 72 IMMERKEHTNLMSLTKKREEKQKALKLLNEVOEHLEERLCRESLADSGWGERSC 131
 Db 61 ilmerreehsklmrtlkkreekqalklmnevqehleeeerlcqvsimgwdeckscl 120
 QY 132 ENNCMRIYTTCPSSWSVKNKIBRFKRIYQFLFPFHEDNEKDLPISEKLIKDAQLTOM 191
 Db 121 esdcmrftttccqsswsmkstiervrklyqflfpfheddekelvgekteedvmlmqi 180
 QY 192 EDVFSQLTVDVNSLFRNSFNFRMQOEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 251
 Db 181 envfsqldvdflymnsfsvfkmgqgefdaafqsfmsdtdmepfyfpafskepaka 240
 QY 252 DLQOCWDIPNFFOLFNCNFSVIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 311
 Db 241 hmqswdipffqfncfslsvygsatvtemikaedlskqdkahggsstcwpvr 300
 QY 312 DRGLCGELDONLSRCFKHEKCKQKQAHLSDECDPVPALHTEDEAIRLVNVSNOQYGOI 371
 Db 301 grglcgepgqnsseciqfharccqkcdylwadcpavpelytkadealelvnlsnqyagv 360
 QY 372 LQMTREKHEDTAYLVEKMRGQFGWSELANOAPETETIIFNSIQVVPRIHEGNISKODETM 431
 Db 361 lqmtqhlhdedtlylmekmregfgwvvelasqtpgseniifsklvvpgvhggnfskqdkm 420
 QY 432 MTDLSILPSSNFTLKIPLESSESSNFYGVVAKALQHEKFEKFTW 477
 Db 421 i-disilpsnftitpleesaeessfismyiaakavqhkhfksw 465

RESULT 6
 AAY30787
 ID AAY30787 standard; Protein; 466 AA.
 XX
 AC AAY30787;
 XX
 DT 23-NOV-1999 (first entry)
 XX
 DE Protein encoded by the guinea pig HKNG1 ortholog gphkng1815.

XX HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
 KW schizophrenia; splice variant.

OS Cavia cobaya.

XX WO9947535-A1.

PN 23-SEP-1999.

PD 16-MAR-1999; 99WO-US05606.

XX 16-MAR-1998; 98US-0078044.

PR 05-JUN-1998; 98US-0088312.

PR 28-OCT-1998; 98US-0106056.

PR 22-JAN-1999; 98US-0236134.

XX (MILL-) MILLENNIUM PHARM INC.

PA (REGC) UNIV CALIFORNIA.

XX Chen H, Freimer NB;

XX WPI; 1999-562047/47.

DR N-PSDB; AAZ10755.

XX New HKNG1 polynucleotides useful in diagnosis and treatment of

PT neuropsychiatric disorders, e.g. bipolar affective disorders and

PT schizophrenia

XX Claim 1; Fig 7A-B; 205pp; English.

CC The present sequence is encoded by guinea pig HKNG1 ortholog gphkng1815.
 CC HKNG1 (Hong Kong new gene 1) is a gene associated with bipolar affective
 CC disorder (BAD). HKNG1 polynucleotides are useful to identify compounds
 CC modulating HKNG1 gene expression or HKNG1 polypeptide expression/
 CC activity. Compounds inhibiting or enhancing HKNG1 gene expression or
 CC activity in individuals can then be administered therapeutically to
 CC treat HKNG1-mediated disorders, especially neuropsychiatric disorders
 CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as
 CC early-onset autosomal dominant myopia. The polynucleotides can be used
 CC in gene therapy techniques to treat such disorders. They are also useful
 CC in diagnosis to identify individuals having, or at risk of developing,
 CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such
 CC mutations especially result in the production of a protein with a
 CC different sequence to the human full-length HKNG1 polypeptide or
 CC splice variant sequences, especially the substitution of a lysine for
 CC a glutamic acid at residue 202 or 184. The polynucleotides are also
 CC useful in gene mapping, to produce probes or primers to identify
 CC similar sequences (e.g. mutants or sequences from different species)
 CC and to produce transgenic animals.
 XX Sequence 466 AA;
 SQ

Query Match 64.7%; Score 1635; DB 20; Length 466;
 Best Local Similarity 66.5%; Pred. No. 2,3e-129;
 Matches 311; Conservative 64; Mismatches 89; Indels 4; Gaps 3;

QY 12 MKPPLLVFVICLLMLKDSHCAPTWKDKTAISENLKSFSEVGEIDAEVYKALTIKQMK 71

Db 1 mkppllmfpvcillwkddchcaptopktdktaisenansfseageidvdvgekvialigkqm 60

QY 72 IMMERKEHTNLMSLTKKREEKQKALKLLNEVOEHLEERLCRESLADSGWGERSC 131

Db 61 immerreehsklmrtlkkreekqalklmnevqehleeeerlcqvsldswdceacl 120

QY 132 ENNCMRIYTTCPSSWSVKNKIBRFKRIYQFLFPFHEDNEKDLPISEKLIKDAQLTOM 191

Db 121 esdcmrftttccqsswsmkstiervrklyqflfpfheddekelvgekteedvmlmqi 179

QY 192 EDVFSQLTVDVNSLFRNSFNFRMQOEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 251

Db 180 ehvfsqldvdflymnsfsvfkmgqgefdaafqsfmsdtdmepfyfpafskepaya 239

QY 252 DLQOCWDIPNFFOLFNCNFSVIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 311

Db 240 daepsalpnvfiqlchlsfsvygsatvtemikaedlskqdkahggsstcwpvr 299

QY 312 DRGLCGELDONLSRCFKHEKCKQKQAHLSDECDPVPALHTEDEAIRLVNVSNOQYGOI 371

Db 300 drgsdglqgnlsdcvnrfrkrcqcdylwadcpavpelytkadealelvnlsnqyagv 359

QY 372 LQMTREKHEDTAYLVEKMRGQFGWSELANOAPETETIIFNSIQVVPRI--HEGNISKODE 429

Db 360 vqmtqghledtlylmekmregfgwvvelasqtpgseniifsklvvpgvhggnfskqdkm 419

QY 430 TMTDLSILPSSNFTLKIPLESSESSNFYGVVAKALQHEKFEKFTW 477

Db 420 tvvps-silpsnftlsspleksagnanfidhvvekvilqhkhfkfw 466

RESULT 7

AAY30788

ID AAY30788 standard; Protein; 450 AA.

XX AAY30788;

XX AAY30788;

DT 23-NOV-1999 (first entry)

XX Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7b.

XX HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;

XX neuropsychiatric disorder; early-onset autosomal dominant myopia;

XX schizophrenia; splice variant.

CC a glutamic acid at residue 202 or 184. The polynucleotides are also
CC useful in gene mapping, to produce probes or primers to identify
CC similar sequences (e.g. mutants or sequences from different species)
CC and to produce transgenic animals.
XX
SQ Sequence 374 AA;

Query Match 50.5%; Score 1277; DB 20; Length 374;
Best Local Similarity 54.5%; Pred. No. 2,6e-99;
Matches 255; Conservative 44; Mismatches 73; Indels 96; Gaps 3;
QY 12 MKPPLLVFVCLLMKSHCAPTWKDKTAISENLKSFSEGEIDAEVKKALTGKQK 71
DB 1 mkpllmfpvcllwkdchcactwkdktaisenansfseageldvdgkvialigkqm 60
QY 72 IMMERKEHTNLMSTLKKCKREEKQKALKLNVEQHEERLCRESLADSWGECRSL 131
DB 61 immerreeehsklmktlkkckeeqkaiklmnevheleeeslcqvsldswdecraci 120
QY 132 ENNCMRIYTTQPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTOM 191
DB 121 esncmrfdttcqpawssvkn----- 140
QY 192 EDVFSQLTVDVNSLNFNSFNFRMQQEPDQTFQSHFISDTDLTEPYFFPAFKEPMTKA 251
DB 141 -----mepayra 147
QY 252 DLEQCDIPNFQFCNFSVSVESVETITKMLKATEDLPKDKAPDHGGLISKMLPGQ 311
DB 148 daepsWalpnvqllcnlsfsvyqsvsekittlratedppkqdkdnggppskilpeq 207
QY 312 DRGLCGELDQNLSCRFKHEKCKQKQAHLSDECDPVPALHTEDEAIRLVNVSNOQYQGI 371
DB 208 drgsdgklgnlsdcvnrkrckqcdqylsdcdpnvpelyrelnealrlvrsnqyqdv 267
QY 372 LQMTKRKHEDTAYLVKMRGQFGWVSELANQAPETETIFNSIQVVPRI--HEGNISKODE 429
DB 268 vqmtqghledtllmekmrqfgwvseIayqspgaedifnpvkvmaIsahegnssddqd 327
QY 430 TMWTDLSILPSSNFTLKIPLESSESSNFYGVVAKALQHFKEHFKTW 477
DB 328 tvvps-silpssnftlsspleksagnanfidhvvekvlgfhfkehftw 374

RESULT 9

AA37090
ID AAY30790 standard; Protein; 373 AA.
AC AAY30790;
XX
XX 23-NOV-1999 (first entry)
DE Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7d.
XX
XX HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW schizophrenia; splice variant.
XX
OS Cavia cobaya.
PN W09947535-A1.
XX
XX 23-SEP-1999.
XX
XX 16-MAR-1999; 99WO-US05606.
XX
XX 16-MAR-1998; 98US-0078044.
PR 05-JUN-1998; 98US-0088312.
PR 28-OCT-1998; 98US-0106056.
PR 22-JAN-1999; 99US-0236134.
XX
XX (MILL-) MILLENNIUM PHARM INC.

(REGC) UNIV CALIFORNIA.

Chen H, Freimer NB;

WPI; 1999-562047/47.

N-PSDB; AA210758.

XX
PT New HKNG1 polynucleotides useful in diagnosis and treatment of
PT neuropsychiatric disorders, e.g. bipolar affective disorders and
XX schizophrenia -
PS Claim 1; Fig 10A-B; 205pp; English.

XX The present sequence is encoded by a guinea pig HKNG1 ortholog
CC gphkng1815 splice variant gphkng7d. HKNG1 (Hong Kong new gene 1)
CC is a gene associated with bipolar affective disorder (BAD). HKNG1
CC polynucleotides are useful to identify compounds modulating HKNG1
CC gene expression or HKNG1 polypeptide expression/activity. Compounds
CC inhibiting or enhancing HKNG1 gene expression or activity in
CC individuals can then be administered therapeutically to treat
CC HKNG1-mediated disorders, especially neuropsychiatric disorders
CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as
CC in gene therapy techniques to treat such disorders. They are also useful
CC in diagnosis to identify individuals having, or at risk of developing,
CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such
CC mutations especially result in the production of a protein with a
CC different sequence to the human full-length HKNG1 polypeptide or
CC splice variant sequences, especially the substitution of a lysine for
CC a glutamic acid at residue 202 or 184. The polynucleotides are also
CC useful in gene mapping, to produce probes or primers to identify
CC similar sequences (e.g. mutants or sequences from different species)
CC and to produce transgenic animals.

SQ Sequence 373 AA;

Query Match 50.3%; Score 1272.5; DB 20; Length 373;
Best Local Similarity 54.5%; Pred. No. 6.3e-99;
Matches 255; Conservative 45; Mismatches 71; Indels 97; Gaps 4;

QY 12 MKPPLLVFVCLLMKSHCAPTWKDKTAISENLKSFSEGEIDAEVKKALTGKQK 71
DB 1 mkpllmfpvcllwkdchcactwkdktaisenansfseageldvdgkvialigkqm 60
QY 72 IMMERKEHTNLMSTLKKCKREEKQKALKLNVEQHEERLCRESLADSWGECRSL 131
DB 61 immerreeehsklmktlkkckeeqkaiklmnevheleeeslcqvsldswdecraci 120
QY 132 ENNCMRIYTTQPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTOM 191
DB 121 esncmrfdttcqpawssvkn----- 140
QY 192 EDVFSQLTVDVNSLNFNSFNFRMQQEPDQTFQSHFISDTDLTEPYFFPAFKEPMTKA 251
DB 141 -----mpay-----ra 146
QY 252 DLEQCDIPNFQFCNFSVSVESVETITKMLKATEDLPKDKAPDHGGLISKMLPGQ 311
DB 147 daepsWalpnvqllcnlsfsvyqsvsekittlratedppkqdkdnggppskilpeq 206
QY 312 DRGLCGELDQNLSCRFKHEKCKQKQAHLSDECDPVPALHTEDEAIRLVNVSNOQYQGI 371
DB 207 drgsdgklgnlsdcvnrkrckqcdqylsdcdpnvpelyrelnealrlvrsnqyqdv 266
QY 372 LQMTKRKHEDTAYLVKMRGQFGWVSELANQAPETETIFNSIQVVPRI--HEGNISKODE 429
DB 267 vqmtqghledtllmekmrqfgwvseIayqspgaedifnpvkvmaIsahegnssddqd 326
QY 430 TMWTDLSILPSSNFTLKIPLESSESSNFYGVVAKALQHFKEHFKTW 477
DB 327 tvvps-silpssnftlsspleksagnanfidhvvekvlgfhfkehftw 373


```

RESULT 10
RAY30794
ID AAY30794 standard; Protein; 521 AA.
XX
AC AAY30794;
XX
DT 23-NOV-1999 (first entry)
XX
DE Amino acid sequence of human HKNG1 splice variant HKNG1-delta7.
XX
KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW schizophrenia; splice variant.
XX
OS Homo sapiens.
XX
PN WO9947535-A1.
XX
PD 23-SEP-1999.
XX
PF 16-MAR-1999; 99WO-US05606.
XX
PR 16-MAR-1998; 98US-0078044.
PR 05-JUN-1998; 98US-0088312.
PR 28-OCT-1998; 98US-0106056.
PR 22-JAN-1999; 99US-0236134.
XX
PA (MILL-) MILLENNIUM PHARM INC.
PA (REGC ) UNIV CALIFORNIA.
XX
PI Chen H, Frelmer NB;
XX
WPI: 1999-562047/47.
DR N-PSDB; AAZ10762.
XX
PT New HKNG1 polynucleotides useful in diagnosis and treatment of
PT neuropsychiatric disorders, e.g. bipolar affective disorders and
PT schizophrenia.
XX
PS Claim 1: Fig 18A-B; 205pp; English.
XX
CC The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)
CC splice variant. HKNG1 is a gene associated with bipolar affective
CC disorder (BAD). HKNG1 polynucleotides are useful to identify compounds
CC modulating HKNG1 gene expression or HKNG1 polypeptide expression/
CC activity. Compounds inhibiting or enhancing HKNG1 gene expression or
CC activity in individuals can then be administered therapeutically to
CC treat HKNG1-mediated disorders, especially neuropsychiatric disorders
CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as
CC early-onset autosomal dominant myopia. The polynucleotides can be used
CC in gene therapy techniques to treat such disorders. They are also useful
CC in diagnosis to identify individuals having, or at risk of developing,
CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such
CC mutations especially result in the production of a protein with a
CC different sequence to the human full-length HKNG1 polypeptide or
CC splice variant sequences, especially the substitution of a lysine for
CC a glutamic acid at residue 202 or 184. The polynucleotides are also
CC useful in gene mapping, to produce probes or primers to identify
CC similar sequences (e.g. mutants or sequences from different species)
CC and to produce transgenic animals.
XX
SQ Sequence 521 AA:

Query Match 30.9%; Score 780.5; DB 20; Length 521;
Best Local Similarity 73.9%; Pred. No. 3.2e-57;
Matches 156; Conservative 8; Mismatches 16; Indels 31; Gaps 3;

QY 7 SNGNKKPPLLVITVCLLWLLKDSHCAPTWKDKTAISENLKSPSEVGEIDADEVKALTG 66
:|||||
Db 118 nngnmkpllvifvclllwikdshcptwkdktaisenlksfsevgideadeevkaltg 177

```

```

QY 67 IKQKIMMERKEKHTNLMSTLKKCREEKQALKNVQEHLEERLCRESLADSWGE 126
|||||
Db 178 ikqmkimmerkehtnlmstlkkcreekqalkllnevqehleerlcresladswe 237
|||||
QY 127 CRSCLENNCRRIYTTCCPSSSSVKNKI--BRFRKIYQFLFPPRHEDNEKDLPISEKLEIK 184
|||||
Db 238 crsclenncmriyttccpssswsvknklltteafrcy----- 274
|||||
QY 185 DAQLTQMEDVFESQLT---VDVNSLFRNSFNV 212
|:|:|
Db 275 ---lgrtedcvgnltricqdvsnfmknakv 302
|:|:|

RESULT 11
AAU28048
ID AAU28048 standard; Protein; 449 AA.
XX
AC AAU28048;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secretory protein, Seq ID No 217.
XX
KW Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;
KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;
KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;
KW amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;
KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;
KW gut protection; lung; liver fibrosis; immune deficiency; infection;
KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;
KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;
KW fertility; analgesic; pain; antigen.
XX
OS Homo sapiens.
XX
PN WO200166689-A2.
XX
PD 13-SEP-2001.
XX
PF 05-MAR-2001; 2001WO-US04942.
XX
PR 07-MAR-2000; 2000US-0519705.
PR 19-MAY-2000; 2000US-0574454.
PR 17-JUN-2000; 2000US-0596193.
PR 14-JUL-2000; 2000US-0616847.
PR 19-SEP-2000; 2000US-0665363.
PR 20-OCT-2000; 2000US-0693267.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;
PI Zhao QH, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;
XX
WPI: 2001-589934/66.
DR N-PSDB; AAS44948.
XX
PT Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis and treatment of
PT cancer, neurological, inflammatory, and autoimmune disorders -
XX
Example 3; SEQ ID No 217; 107pp; English.
XX
CC The invention relates to novel isolated human secreted polypeptides (I)
CC and polynucleotides (II). (I) and (II) are useful for treating
CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,
CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is
CC involved in increasing haematopoiesis, stem cell survival, bone growth
CC and remodeling. (I), (II) and modulators of (II) are useful for
CC prophylaxis or treatment of one or more cancers. (II) is also useful for
CC creating transgenic animals useful for studying the in vivo activities of
CC the polypeptide as well as for studying modulators of the polypeptides.
CC (I) induces the proliferation of neural cells and regeneration of nerve
CC and brain tissue and is useful for the treatment of central and

```

CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,
 CC Parkinson's disease, Huntington's disease, and amyotrophic lateral
 CC sclerosis. In addition, (I) is involved in chemotactic or chemokinetic
 CC activity, regulation of haematopoiesis and is useful for treating myeloid
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
 CC tissue growth, and in tissue repair, healing of burns, incisions,
 CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative
 CC disorders, or periodontal disease. Furthermore, (I) is also useful for
 CC gut protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues, various immune deficiencies and
 CC disorders including severe combined immunodeficiency (SCID), bacterial or
 CC fungal infections, autoimmune disorders e.g. multiple sclerosis,
 CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic
 CC reactions and conditions, such as asthma or other respiratory problems.
 CC In addition, (I) affects biorhythms or circadian cycles of rhythms,
 CC fertility, metabolism, catabolism, anabolism, storage or elimination of
 CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides
 CC analgesic effects or other pain reducing effects, immunoglobulin like
 CC activity and can act as an antigen in a vaccine composition to raise an
 CC immune response. AAU28020-AAU28395 represent novel human secreted protein
 CC amino acid sequences of the invention.
 XX
 SQ Sequence 449 AA;

Query Match 17.2%; Score 434; DB 22; Length 449;
 Best Local Similarity 25.8%; Pred. No. 4.3e-28;
 Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

QY 12 MKPPLIVFVCLLWKDSHCAPTWK-----DKTAISENLKSFSEVGEIDAEVKKALT 65
 DB 1 mmktllllfvgl-----twesgqvlgdqtdsnelqemsgskynkeiqnavn 51
 QY 66 GIKQIMMERKEKHTNLMSTLKKCRKQKALKLLNEVQHELEERLCRESLADSG 125
 DB 52 gvqkltlektneerktllsneeakkkedalnetresetkikelpgvcnmmalwe 111
 QY 126 ECRSLNENCMRIYT-TCOPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIK 184
 DB 112 eckpckqtkmkyarvcrgsglvgqlleefngspfyfwmngdr-----idsllen 165
 QY 185 DAQLTQMEDV----FSQLTVDVNSLNRSNFVRQMQQEDQTFQSHFTS-DTDLTPEYF 239
 DB 166 drqqtthmldvmqdhfrassiidelfqdrf-----ftrepqdy--hylpfslphrphf 218
 QY 240 FPAFSKEPMTKADLE-QCWDIPNFFQFCNFSVIYESVSETITKMLKAIEDLPKQDKAP 298
 DB 219 f--fpksrivrslmpsfpyepnfhmfqpflemlhea-----qqamdihfhsfaf 267
 QY 299 DHGGLISKMLPGDRLGCGELDQNLRCFKFHEKCKQKQAHLSDEC----PDVPALHTEL 354
 DB 268 qhpptefiregddrtvcvceirhnstgclrmkdqcdkreilsvdcstnnpqaklrrel 327
 QY 355 DEAIRLVNVSQYGOILQWTRKHELDYALVEKMRQCGVGSSELANQAPETETPNSTO 414
 DB 328 deslgvaerltrkynellksyqwkmlntsslleqlneqfnwvrsrlantlqgedqyylyvt 387
 QY 415 VVPRIHGNIISKODETMMTDLSTLPSNFTLKIPLPESAESSNFGYVYVAKALQHF-KEH 473
 DB 388 tvas-htsdsdvpsgvtevvvklfdspdtvtvpvevsknpkfmctvaekalqeyrkhh 446

RESULT 12
 AAR11704
 ID AAR11704 standard; Protein; 448 AA.
 XX
 AC AAR11704;
 XX
 DT 20-JUN-1991 (first entry)
 XX
 DE Cytolysis Inhibitor.
 XX

KW cytolysis inhibitor; perforin; immunological effector molecule;
 KW infertility.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Peptide 1..21 /label= signal peptide
 FT Protein 22..226 /label= A-chain
 FT Protein 227..448 /label= B-chain
 FT
 XX DE3933850-A.
 PN
 XX 18-APR-1991.
 PD
 XX 06-OCT-1989; 89DE-3933850.
 PF
 XX 06-OCT-1989; 89DE-3933850.
 PR
 XX (SCHD) SCHERING AG.
 PA
 XX Tschopp J, Jenne D;
 XX WPI; 1991-118338/17.
 DR
 XX
 XX DNA sequence coding for cytolysis inhibitor - is strong inhibitor
 PT of terminal complement protein, eg perforin secreted by killer
 PT cells
 PT
 XX
 PS Claim 13; Page 9; 15pp; German.
 CC
 CC This cytolysis inhibitor is encoded by a 1.7kb BamHI-KpnI fragment
 CC isolated from a liver-specific cDNA library. It is a blood plasma
 CC component that inhibits immunological effector molecules. It is used
 CC for systemic or local treatment of inflammatory or autoimmune
 CC diseases mediated by complement or killer cells. It can also be
 CC used for detoxification of membrane-active and cytolytic proteins
 CC released by bacteria, fungi and insect venoms. The protein has a
 CC further use in the treatment of infertility caused by deficiency of
 CC cytolysis inhibitor. Monoclonal antibodies directed against the
 CC cytolysis inhibitor are also covered by the invention. They are used
 CC to isolate or quantify the natural protein in human plasma.
 CC See also AAQ11501 and AAQ11502.
 XX
 SQ Sequence 448 AA;

Query Match 17.1%; Score 432; DB 12; Length 448;
 Best Local Similarity 25.8%; Pred. No. 6.3e-28;
 Matches 123; Conservative 96; Mismatches 205; Indels 52; Gaps 14;

QY 16 LLVFIVCLLWKDSHCAPTWK-----DKTAISENLKSFSEVGEIDAEVKKALTGIKQ 69
 DB 4 lllfvgl-----twesgqvlgdqtdsnelqemsgskynkeiqnavngvk 54
 QY 70 MKIMMERKEKHTNLMSTLKKCRKQKALKLLNEVQHELEERLCRESLADSGECS 129
 DB 55 iktlektneerktllsneeakkkedalnetresetkikelpgvcnmmalweeekp 114
 QY 130 GLENENCMRIYT-TCOPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAOL 188
 DB 115 clkqtkmkyarvcrgsglvgqlleefngspfyfwmngdr-----idsllenrdq 168
 QY 189 TQMEDV----FSQLTVDVNSLNRSNFVRQMQQEDQTFQSHFTS-DTDLTPEYFFAF 243
 DB 169 thmldvmqdhfrassiidelfqdrf-----ftrepqdy--hylpfslphrphff--f 219
 QY 244 SKEPMTKADLE-QCWDIPNFFQFCNFSVIYESVSETITKMLKAIEDLPKQDKAPDHG 302
 DB 220 pksrivrslmpsfpyepnfhmfqpflemlhea-----qqamdihfhsfafqhpp 270

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:52:30 ; Search time 23.96 Seconds
(without alignments)
486.269 Million cell updates/sec

Title: US-09-722-544A-4MOD
Perfect score: 2529
Sequence: 1 MFTWYSNGNMKPLLVI.....FTGVYVAKALQHKRHKWT 477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5b_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCUTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2483	98.2	495	4	US-09-268-992-2
2	2460	97.3	466	4	US-09-268-992-64
3	2373	93.8	477	4	US-09-268-992-4
4	2346	92.8	446	4	US-09-268-992-51
5	1751.5	69.3	465	4	US-09-268-992-49
6	1635	64.7	466	4	US-09-268-992-39
7	1554	61.4	450	4	US-09-268-992-41
8	1277	50.5	374	4	US-09-268-992-43
9	1272.5	50.3	373	4	US-09-268-992-45
10	776	30.7	208	4	US-09-268-992-67
11	149	5.9	1388	2	US-08-685-576-1
12	145	5.7	1388	2	US-08-685-576-4
13	130.5	5.2	2482	1	US-08-328-254-6
14	130	5.1	3248	1	US-08-353-700-1
15	130	5.1	3248	5	PCT-US95-16216-1
16	126.5	5.0	1354	3	US-08-685-871-2
17	122	4.8	1886	4	US-08-938-105-3
18	121	4.8	828	2	US-08-993-228-21
19	120.5	4.8	1618	1	US-07-853-913-4
20	119.5	4.7	1786	4	US-08-973-462-8
21	119.5	4.7	3111	2	US-08-460-309-4
22	119.5	4.7	3111	2	US-08-125-077-4
23	118.5	4.7	816	2	US-08-533-306A-6
24	118.5	4.7	816	2	US-08-742-923A-6
25	117	4.6	1098	4	US-08-923-992A-8
26	117	4.6	1104	4	US-08-923-992A-4
27	117	4.6	1939	4	US-09-310-187A-1

28	117	4.6	2052	3	US-09-045-201A-2	Sequence 2, Appli
29	116.5	4.6	885	2	US-08-533-306A-4	Sequence 4, Appli
30	116.5	4.6	885	2	US-08-742-923A-4	Sequence 4, Appli
31	116.5	4.6	1057	4	US-09-541-782-10	Sequence 10, Appli
32	116	4.6	1164	4	US-08-923-992A-2	Sequence 2, Appli
33	114	4.5	435	2	US-08-531-439B-4	Sequence 4, Appli
34	114	4.5	1128	4	US-08-923-992A-6	Sequence 6, Appli
35	114	4.5	2154	2	US-08-841-349-4	Sequence 18, Appli
36	113.5	4.5	452	2	US-08-686-599A-18	Sequence 5, Appli
37	113.5	4.5	493	2	US-08-686-599A-5	Sequence 16, Appli
38	113.5	4.5	493	2	US-08-686-599A-16	Sequence 17, Appli
39	111.5	4.4	467	2	US-08-686-599A-17	Sequence 8, Appli
40	111.5	4.4	1066	4	US-09-541-782-8	Sequence 4, Appli
41	109	4.3	976	4	US-09-104-324B-4	Sequence 10, Appli
42	108	4.3	2285	4	US-09-308-375-2	Sequence 7, Appli
43	107.5	4.3	1038	4	US-09-541-782-4	Sequence 4, Appli
44	107	4.2	1164	4	US-08-923-992A-10	Sequence 4, Appli
45	107	4.2	1497	1	US-08-623-679-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-268-992-2
; Sequence 2, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freilmer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-2

Query Match	98.2%;	Score	2483;	DB	4;	Length	495;
Best Local Similarity	99.6%;	Pred. No.	1.9e-230;				
Matches	469;	Conservative	2;	Mismatches	0;	Indels	0;
0;							
QY	7	SNNGNMKPLLVFIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDAEVVKKALTG	66				
Db	25	NNSGNMCKPLLVFIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDAEVVKKALTG	84				
QY	67	IKQKIMMERKEKHTNLMSTLKKCREKQKALKNLNEVQHLFEERLCRESLADSWGE	126				
Db	85	IKQKIMMERKEKHTNLMSTLKKCREKQKALKNLNEVQHLFEERLCRESLADSWGE	144				
QY	127	CRSLENNCMRIYTCOPSSSVKNKIERFERFKYQFLFPFHEDNEKDLPISEKLIKDA	186				
Db	145	CRSLENNCMRIYTCOPSSSVKNKIERFERFKYQFLFPFHEDNEKDLPISEKLIKDA	204				
QY	187	QLTQMEDVFSQLTVDVNSLNFNRNFVFMQOEFDQTFQSHFISDTLTETPEYFPAPFSKE	246				
Db	205	QLTQMEDVFSQLTVDVNSLNFNRNFVFMQOEFDQTFQSHFISDTLTETPEYFPAPFSKE	264				
QY	247	PMTKADLEQCHWDINFFQLFCNFSVSIYESVETITKMLKAIEDLPKODKAPDHGGLISK	306				

Db 265 PWTADLEQCHDIPNFOLFNFVSIVSEYITTKMLKAIEDLPKQDAPDRGGLISK 324
Qy 307 MLPQDRLGCLGELDONLSRCFKFHEKQKQCAHLSEDCPDVPALHTELDEAIRLVNSNQ 366
Db 325 MLPQDRLGCLGELDONLSRCFKFHEKQKQCAHLSEDCPDVPALHTELDEAIRLVNSNQ 384
Qy 367 QYQILQWTRKHLEDATYLVKMRGQFCGWSELANQAPETETIIFNSIQVVPRIHEGNISK 426
Db 385 QYQILQWTRKHLEDATYLVKMRGQFCGWSELANQAPETETIIFNSIQVVPRIHEGNISK 444
Qy 427 QDETMTDLSILPSSNFTLKIPLESASSNFYGVVAKALQHKEHFKTW 477
Db 445 QDETMTDLSILPSSNFTLKIPLESASSNFYGVVAKALQHKEHFKTW 495

RESULT 2
US-09-268-992-64
; Sequence 64, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER FILING DATE: 1999-01-22
; EARLIER FILING DATE: 1998-10-28
; EARLIER FILING DATE: 1998-06-05
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-64

Query Match 97.3%; Score 2460; DB 4; Length 466;
Best Local Similarity 99.8%; Pred. No. 2.8e-228;
Matches 465; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 12 MKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEEVKKALTGKQMK 71
Db 1 MKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEEVKKALTGKQMK 60
Qy 72 IMMERKEHTNLMSLTKKREKQKALLLNEVQHLSEERLCRESLADSGECSRL 131
Db 61 IMMERKEHTNLMSLTKKREKQKALLLNEVQHLSEERLCRESLADSGECSRL 120
Qy 132 ENNCMRIYTTCPQSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQM 191
Db 121 ENNCMRIYTTCPQSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQM 180
Qy 192 EDVFSQTLVDVNSLFNRSFNVFQMOQEPDQTFQSHFISDTDLTEPYFPFAPSKPEMTKA 251
Db 181 EDVFSQTLVDVNSLFNRSFNVFQMOQEPDQTFQSHFISDTDLTEPYFPFAPSKPEMTKA 240
Qy 252 DLEQCDWIPNFOLFNFVSIVSEYITTKMLKAIEDLPKQDAPDRGGLISKMLPGQ 311
Db 241 DLEQCDWIPNFOLFNFVSIVSEYITTKMLKAIEDLPKQDAPDRGGLISKMLPGQ 300
Qy 312 DRGLCGELDONLSRCFKFHEKQKQCAHLSEDCPDVPALHTELDEAIRLVNSNQYQGI 371
Db 301 DRGLCGELDONLSRCFKFHEKQKQCAHLSEDCPDVPALHTELDEAIRLVNSNQYQGI 360

Qy 372 LQWTRKHLEDATYLVKMRGQFCGWSELANQAPETETIIFNSIQVVPRIHEGNISKODETM 431
Db 361 LQWTRKHLEDATYLVKMRGQFCGWSELANQAPETETIIFNSIQVVPRIHEGNISKODETM 420
Qy 432 MTDLISILPSSNFTLKIPLESASSNFYGVVAKALQHKEHFKTW 477
Db 421 MTDLISILPSSNFTLKIPLESASSNFYGVVAKALQHKEHFKTW 466

RESULT 3
US-09-268-992-4
; Sequence 4, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER FILING DATE: 1999-01-22
; EARLIER FILING DATE: 1998-10-28
; EARLIER FILING DATE: 1998-06-05
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-4

Query Match 93.8%; Score 2373; DB 4; Length 477;
Best Local Similarity 91.3%; Pred. No. 6.7e-220;
Matches 454; Conservative 2; Mismatches 1; Indels 40; Gaps 2;
Qy 1 MRTWDYSNGNKKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEEV 60
Db 1 MRTWDYSNGNKKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEEV 60
Qy 61 KKALTGKQKIMMERKEK-----EHTNLMSTLKKREKQKQALK 100
Db 61 KKALTGKQKIMMERKEKQAQAPETETIIFNSIQVVPRIEHTNLMSTLKKREKQKQALK 120
Qy 101 LLNEVQHLSEERLCRESLADSGECSRLNENCMRIYTTCPQSWSSVKNKIERFRKI 160
Db 121 LLNEVQHLSEERLCRESLADSGECSRLNENCMRIYTTCPQSWSSVKNKIERFRKI 180
Qy 161 YQFLFPFHEDNEKDLPISEKLIKDAQLTQMEDVFSQTLVDVNSLFNRSFNVFQMOQEF 220
Db 181 YQFLFPFHEDNEKDLPISEKLIKDAQLTQMEDVFSQTLVDVNSLFNRSFNVFQMOQEF 240
Qy 221 DQTFQSHFISDTDLTEPYFPFAPSKPEMTKADLEOCWDIPNFOLFNFVSIVSEVSET 280
Db 241 DQTFQSHFISDTDLTEPYFPFAPSKPEMTKADLEOCWDIPNFOLFNFVSIVSEVSET 300
Qy 281 ITKMLKAIEDLPKQDAPDRGGLISKMLPGQDRGLCGELDONLSRCFKFHEKQKQCAHL 340
Db 301 ITKMLKAIEDLPKQDAPDRGGLISKMLPGQDRGLCGELDONLSRCFKFHEKQKQCAHL 360
Qy 341 SEDCPDVPALHTELDEAIRLVNSNQYQGIQWTRKHLEDATYLVKMRGQFCGWSEL 400
Db 361 SEDCPDVPALHTELDEAIRLVNSNQYQGIQWTRKHLEDATYLVKMRGQFCGWSEL 419
Qy 401 NOAPETETIIFNSIQVVPRIHEGNISKODETMTDLSILPSSNFTLKIPLESASSNF 460
Db 420 -----HEGNISKODETMTDLSILPSSNFTLKIPLESASSNF 460


```
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Cavia sp.
US-09-268-992-39
```

```
Query Match 64.7%; Score 1635; DB 4; Length 466;
Best Local Similarity 66.58; Pred. No. 6.4e-149;
Matches 311; Conservative 64; Mismatches 89; Indels 4; Gaps 3;

QY 12 MKPPLLVIYVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVKKALTGKQK 71
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKLPLLMFPVCLLWLDKCHCAPTWKDKTAISENANFSFEAGEIDVDGEVKIALIGK 60
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 72 IMMERKEKHTNLMSTLKKCKEKEQKALKLLNEVQEHLEERLCRESLADSWGECR 131
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 IMMERREEHSLKMTLKKCKEKEQKALKLLNEVHEHELEESLCQVSLADSWDECR 120
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 132 ENNCRIYTTCPSSVSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLEKDA 191
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 ENSCMREFTTQCPANSSVKNVQEFKRIYQFLPLQE-NDRSGPVSKGVTEADA 179
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 192 EDVFSQLTVDVNSLNFNSFVFRMQOEFQDTQSFHSFISDTEPYFFPAFSKE 251
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 180 EHVFSQLSADVTSLFNRSLYVFKQLRRFEQAFQSYFTSGDVTPEFFPSLSKE 239
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 252 DLEQCWDIPNFFQFCNFSVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLIS 311
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 240 DAEPSWAI PNVPVQLCNLSFSVSIQVSEKLTITLRATEDPPKQDKDSNOGG 299
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 312 DRGLCGELDNLSRCFKFHEKQKQAHLSDECDPVPALHTELDEAIRLVNSNQY 371
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 DRGSDGKLGONLSDCNVFRKRCQKQDYLSDDCPNVPPELYRELNEALRLVSR 359
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 372 LQWTRKHLEDYALVEKMRGQFVWSELANOAPETETIFNSIQVVPRI--HEGN 429
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 360 VQMTQYHLEDYTLMEKMRQFVWSELAYQSPGAEDIFNPVKVMVALSAHEGNS 419
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 430 TMMTDLSTLSPSSNFTLKIPLEESAESNFIQVYVAKALQHFKEHFKTW 477
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 420 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVQLQHFKEHFKTW 466
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
RESULT 7
US-09-268-992-41
; Sequence 41, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
```

```
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Cavia sp.
US-09-268-992-41

Query Match 61.4%; Score 1554; DB 4; Length 450;
Best Local Similarity 63.7%; Pred. No. 3.8e-141;
Matches 298; Conservative 63; Mismatches 87; Indels 20; Gaps 3;

QY 12 MKPPLLVIYVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVKKALTGK 71
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKLPLLMFPVCLLWLDKCHCAPTWKDKTAISENANFSFEAGEIDVDGEVKIALIGK 60
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 72 IMMERKEKHTNLMSTLKKCKEKEQKALKLLNEVQEHLEERLCRESLADSWGECR 131
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 IMMERREEHSLKMTLKKCKEKEQKALKLLNEVHEHELEESLCQVSLADSWDECR 120
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 132 ENNCRIYTTCPSSVSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLEKDA 191
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 ENSCMREFTTQCPANSSVKNV-----ENDRSGPVSKGVTEADAQVSHI 163
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 192 EDVFSQLTVDVNSLNFNSFVFRMQOEFQDTQSFHSFISDTEPYFFPAFSKE 251
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 164 EHVFSQLSADVTSLFNRSLYVFKQLRRFEQAFQSYFTSGDVTPEFFPSLSKE 223
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 252 DLEQCWDIPNFFQFCNFSVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLIS 311
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 224 DAEPSWAI PNVPVQLCNLSFSVSIQVSEKLTITLRATEDPPKQDKDSNOGG 283
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 312 DRGLCGELDNLSRCFKFHEKQKQAHLSDECDPVPALHTELDEAIRLVNSNQY 371
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 284 DRGSDGKLGONLSDCNVFRKRCQKQDYLSDDCPNVPPELYRELNEALRLVSR 343
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 372 LQWTRKHLEDYALVEKMRGQFVWSELANOAPETETIFNSIQVVPRI--HEGN 429
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 344 VQMTQYHLEDYTLMEKMRQFVWSELAYQSPGAEDIFNPVKVMVALSAHEGNS 403
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 430 TMMTDLSTLSPSSNFTLKIPLEESAESNFIQVYVAKALQHFKEHFKTW 477
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 404 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVQLQHFKEHFKTW 450
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
RESULT 8
US-09-268-992-43
; Sequence 43, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
```

; SEQ ID NO 43
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Cavia sp.
US-09-268-992-43

```
Query Match          50.5%; Score 1277; DB 4; Length 374;
Best Local Similarity 54.5%; Pred. No. 1.3e-114;
Matches 255; Conservative 44; Mismatches 73; Indels 96; Gaps 3;

QY 12 MKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVVKALTGKQMK 71
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKPLLMFPVCLLWLDKSHCAPTWKDKTAISENANFSFSEAGEIDVDGEVKIALIGIKQMK 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 72 IMMERKEHTNLMSTLKKCKEKEQKALKNLNEVQEHLEERLCRESLADSWGECRSCL 131
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 IMMERREEHSKLMKTLKKCKEKEQKALKNLNEVHEHEERESLCQVSLADSWDECRACL 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 132 ENNCMRIYTTCCPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQM 191
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 ENSCMRFTDTCQPAWSSVKN-----MPAY-----RA 146
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 192 EDVFSQLTVDVNSLFRNSFNVRMQQEFDTQFSHFISDTDLTEPYFFPAFSKEPMTKA 251
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 141 -----MPAY-----RA 146
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 252 DLEQCWDIPNFPOLFCNFSVSIYESVETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 311
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 148 DAEPSSWAIPNVPOLLNLSFSVQSVSEKLTITLRATEDPPKQDKDSNOGGPISKILPEQ 206
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 312 DRGLGCLDQNLSCRCFHEKCKQKQAHLSDCDPPVPAHTELDDEAIRLVNVSNOQYGOI 371
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 208 DRGSDGLQGLNLSDCVNFRRCKQKQDYLSDDCPNVPYELRELNEALRLVSRNSQYDQV 266
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 372 LQMTKRKHLEDYAYLVEKMRGQFGWVSELANOAPETEIENSIOVVPRI--HEGNISKODE 429
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 268 VQMTQYHLEDYAYLVEKMRGQFGWVSELANOAPETEIENSIOVVPRI--HEGNISKODE 429
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 430 TMMTDLSILPSSNFTLKIPLESASSNFIYGVYVAKALQHFKEHFKTW 477
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 328 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVQLQHFKEHFKTW 374
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 9
US-09-268-992-45
; Sequence 45, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Cavia sp.
US-09-268-992-45
```

```
Query Match          30.7%; Score 776; DB 4; Length 208;
Best Local Similarity 98.6%; Pred. No. 8.1e-67;
Matches 145; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 7 SNSGNMKPPLLVFIVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVVKALTG 66
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 57 NNSGNMKPPLLVFIVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVVKALTG 116
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

US-09-268-992-45
```

```
Query Match          50.3%; Score 1272.5; DB 4; Length 373;
Best Local Similarity 54.5%; Pred. No. 3.4e-114;
Matches 255; Conservative 45; Mismatches 71; Indels 97; Gaps 4;

QY 12 MKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVVKALTGKQMK 71
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKPLLMFPVCLLWLDKSHCAPTWKDKTAISENANFSFSEAGEIDVDGEVKIALIGIKQMK 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 72 IMMERKEHTNLMSTLKKCKEKEQKALKNLNEVQEHLEERLCRESLADSWGECRSCL 131
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 IMMERREEHSKLMKTLKKCKEKEQKALKNLNEVHEHEERESLCQVSLADSWDECRACL 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 132 ENNCMRIYTTCCPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQM 191
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 ENSCMRFTDTCQPAWSSVKN-----MPAY-----RA 146
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 192 EDVFSQLTVDVNSLFRNSFNVRMQQEFDTQFSHFISDTDLTEPYFFPAFSKEPMTKA 251
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 141 -----MPAY-----RA 146
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 252 DLEQCWDIPNFPOLFCNFSVSIYESVETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 311
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 147 DAEPSSWAIPNVPOLLNLSFSVQSVSEKLTITLRATEDPPKQDKDSNOGGPISKILPEQ 206
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 312 DRGLGCLDQNLSCRCFHEKCKQKQAHLSDCDPPVPAHTELDDEAIRLVNVSNOQYGOI 371
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 207 DRGSDGLQGLNLSDCVNFRRCKQKQDYLSDDCPNVPYELRELNEALRLVSRNSQYDQV 266
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 372 LQMTKRKHLEDYAYLVEKMRGQFGWVSELANOAPETEIENSIOVVPRI--HEGNISKODE 429
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 267 VQMTQYHLEDYAYLVEKMRGQFGWVSELANOAPETEIENSIOVVPRI--HEGNISKODE 429
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 430 TMMTDLSILPSSNFTLKIPLESASSNFIYGVYVAKALQHFKEHFKTW 477
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 327 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVQLQHFKEHFKTW 373
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 10
US-09-268-992-67
; Sequence 67, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-67
```

```
Query Match          30.7%; Score 776; DB 4; Length 208;
Best Local Similarity 98.6%; Pred. No. 8.1e-67;
Matches 145; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 7 SNSGNMKPPLLVFIVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVVKALTG 66
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 57 NNSGNMKPPLLVFIVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVVKALTG 116
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

US-09-268-992-67
```

QY 67 IKQKIMMERKEHTNLMSTLKKREKQKALKLLNEVQHLBEERLCRESLADSWGE 126
DB 117 IKQKIMMERKEHTNLMSTLKKREKQKALKLLNEVQHLBEERLCRESLADSWGE 176

QY 127 CRSCLENNCMRIYTCQPSWSSVKNKI 153
DB 177 CRSCLENNCMRIYTCQPSWSSVKNKL 203

RESULT 11
US-08-685-576-1
; Sequence 1, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Koza
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-325129
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136

INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-685-576-1

Query Match 5.9%; Score 149; DB 2; Length 1388;
Best Local Similarity 16.7%; Pred. No. 3e-05;
Matches 95; Conservative 96; Mismatches 166; Indels 212; Gaps 20;

QY 36 KDKTAISENLKSFSEVEIDADEVKALTGKIKOMKI-----MMERKEKHTNLM 86
DB 780 KQKDVNLNEDVNRNL-----LKIEQETQKRCULTQNDLRMTQOVNLTLMKSEKOLKQENHLL 835
QY 87 TLK-----KCREKQKALKLLNEVQHELEEE-----RLCR 117

DB 836 EMKMSLEKQNAELRKERQDADGQMKELOQLEAEQVFTSLYKTQVRELKECECEKTKLCK 895
QY 118 E-----SLADSGEGRSCLENNCMRIYTCQPSWSSVKNKIIFRFRKIYQFLFPFHEDN 171
DB 896 ELQKKQELQDE-----RDSLAQALETITLKADSEQLARSIAEEQYSDLEKE 942
QY 172 E--KDLPISE-----KLIEKDAQLTQMEDVFSQLTVDVNSLFRNSFNVRMQQEFDO 222
DB 943 KIMKELEIKEMMARHKQELTEKDATIASLEETNRTLTSDVANLANEKEELNNKLKEAQEQ 1002
QY 223 TFQSHFISDLDLTPYFFPAFSKEPMTKADLEQCDWDIPNFFQLFCNFSVSIYESVETIT 282
DB 1003 LSR---LKDEEISAAATKAQFEXOLLTERLK-----TQAVN 1036
QY 283 KMLKAIE-----DLPKQDKAPDGHGLISKMLPGQDRGLCGEL-----DQNLISRCF 327
DB 1037 KLAIMNRKEPVKRGNDTDVRRKEK-----ENRKLHMLKSEREKLTOQMI 1082
QY 328 KFHEKQKQCAHJSEDCPDVPAHTELDIAIRLVNVSNNQYGGILQMTKRHLE-----380
DB 1083 KYQKELNEMOQAIIE-----SQIRIELQMTLDSKSDIEQLRSQALQALHIGLSSSTGSG 1138
QY 381 -----DTAYLVEKMRG-----QEGWVSEL-----399
DB 1139 PGDTEADGGFPESRLEGLWSLPLVRNNTKKKGWVKYVIVSSKKILFYDSEQDKESNPYM 1198
QY 400 -----ANQAPETETIENSIOQVPR-----HEGNTSKQDETMMTDLISILPSSNFT 444
DB 1199 VLDIDKLFHVRPVQTQDVYRADAKEIPRIFOILYANEGESKKEQE-----1243
QY 445 LKIPLEESAESSNFI---GYVVAKALQHF 470
DB 1244 --FPVEPVGEKSNVICHKGHEFIPTLYHF 1270

RESULT 12
US-08-685-576-4
; Sequence 4, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Koza
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-325129
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996

ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16887/843
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1388 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-685-576-4

Query Match 5.7%; Score 145; DB 2; Length 1388;
Best Local Similarity 17.0%; Pred. No. 7.3e-05;
Matches 95; Conservative 96; Mismatches 176; Indels 192; Gaps 20;
QY 36 KKTALSENKSEFSEGEIDAEVYKALTGKIKOMKI-----MMERKEKEHTNLS 86
DB 780 KOKDVLNEDVRNLT-----LKIEQTKRCLTQNDLKMTQOVNTLKSEKOLKQENHLM 835
QY 87 TLK-----KCREKQKALKNLEVOHLEEE-----RLCRESLADSWGEC-----127
DB 836 ENKMNLKQNALRKERDQAGOMKELQDLAEQYFTSLYKTQVRELKECEKTKLKG 895
QY 128 ---RSCLENNCMRYITTCQPSWSSVKNKIERFFRYQFLFPFHEDNE--KOLPIS--179
DB 896 ELQOKKQELQDRDSLAQALETTLTKADSEQLARSIAEQYSDLEKIMKEIKEMMA 955
QY 180 ---KLIKDAQLOMDEVFSQLVNLSFNRS---FNVRQMQQEDQTFQSHFISDT 232
DB 956 RHKQELTKDATIASLEETNTLSVLANLANEKEELNKLKDVQQLSR-----LKDE 1009
QY 233 DLTEPYFPFAPSKPMTKADLEQCDWIPNFFOLFCNFSVSIYESVETITKMLKAIE---289
DB 1010 EISAAAKAQPEKQLLTERTLK-----TQAVNKLAEIMNRKE 1046
QY 290 ---DLPKQKADPHGGLISKMLPGQDRGLCGEL-----DONLSRCFKFHEKQKQ 337
DB 1047 PVKRGNDTDVRKKE-----ENRKLHMLKSEREKLQOMIKYQKELNEMQ 1092
QY 338 AHLSEDCPDVPAHTELDIAIRLVNVSNOQYQGLQWTRKHLF-----DTAY 384
DB 1093 AQIAEE---SQRIELQWTLDSKSDIEQLRSQALHGLDSSSIGSGPGDAEADDGF 1148
QY 385 LVEKMRG-----QFGWVSEL-----A 400
DB 1149 PESLEGWLSLPVRNNTKFGVKKYVIVSSKKILFYDSEQKQSNPYWVLIDIKLPHV 1208
QY 401 NQAPETEIFNSIQVVPRI-----HEGNISKQDETMTDLSILSSNTFLKIPLESAAE 454
DB 1209 RPTQTDVYRADAKEIPRIQILYANEGESKKEQ-----FPVEPVGE 1251
QY 455 SSNFI---GYVAKALQHF 470
DB 1252 KSNYICHGHEFTPLYHF 1270

RESULT 13
US-08-328-254-6
Sequence 6, Application US/08328254
Patent No. 5710022
GENERAL INFORMATION:
APPLICANT: Zhu, Xueliang
APPLICANT: Lee, Wen-Hwa
TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,254
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/141,239
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 1191
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2482 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-328-254-6

Query Match 5.2%; Score 130.5; DB 1; Length 2482;
Best Local Similarity 18.8%; Pred. No. 0.0044;
Matches 99; Conservative 85; Mismatches 165; Indels 177; Gaps 22;
QY 27 KDSHCAPTWKDKTALSENK-----SFSEVGE-----IDADE---EVKALTGKIQ 69
DB 1502 KOLHIAEKLKERENDSLKQKVENLERELQMSQENQELVILDAENSKAEVETLKTQIEE 1561
QY 70 M-----KIMMERKEKEH-----TNLMSTLKKCKREEKQKALKNLEEV 105
DB 1562 MARSLKVFELDLVTLRSEKENLTQIQEKQKQSELQKLLSSFKSLLEKEQA-----EI 1616
QY 106 OEHLEERELCRESLADSWGECRSCLENNC--MRIYTCQPSWS-----SVKNKIERF 156
DB 1617 Q--IKESKTAVMLQNLKELNEVAALCGDQEIWKATEQSLDPPTEEEHQLRNSIEKL 1674
QY 157 FRKIYQFLFPFHEDNEKDLPISEKLEKD-----185
DB 1675 RARL-----EADKQKQVLOOLKESEHDLKGRVENLERELEIARTNOEHALEA 1727
QY 186 -----AQLOMDEVFSQLVNLSFNRSFNVRQMQQEDQTFQSHFISDTDL 234
DB 1728 ENSGVEVETLKAKTEGTSRLGLELDVVTIRSEKENLTNELQKEQER-----ISELEI 1781
QY 235 TEPYFFPAPSKPMTKADLEQCDWIPNFFOLFCNFSVSIYESVETITKMLKA-IEDLPK 293
DB 1782 INSSFENILQKEQKQVQKME-----KSSSTAMEMLQTLKELNE 1820
QY 294 QDKAPDHGGLISKMLPGQDRGLCGELQDNLS---RCFKFHEKQKQAHLSDECDDVPAL 350
DB 1821 RVAALHN-----DQEAACKAKEQNLSSQVECLEL-EKAQLLQ-----1856
QY 351 HTLDEAIRLVNVSNOQYQGLQWTRKHLFDTAYLVEKMRGQFGVWSELANQAPETEIF 410
DB 1857 ---LDEAKNNYIVLQSSVNGLIQ-----EVEDGKQKLEKDEE---ISRLKNOIQOEQLV 1906
QY 411 NSIQVVPRIH-----EGNISQDETMTD--LSILPSSNFTFLKIPLE 450
DB 1907 SKLSQVEGEHQLWKQNLRLNLTVELEQKIQVLOSKNASLQDTLE 1952

Qy 147 SSVNKIERFR--KIYQF-LFPFHEDNEKDLPISEKLIKDAQLTOMEDVFS----- 196
Db 2283 --LKTQIEENARSLKIFELDLVTLRSEKEN---LTKQIQEKQQLSSELDKLLSSFKSLLE 2337
Qy 197 --OLTVDVNSLFRNSFNFRQOQEFDTQFQSHFISDTDL---TEPYFFPAFSKEPMWK 250
Db 2338 EKEQREIOIREEKTAVENTLONOLKELNEAVAA-LCGDQEIIMRATEQSLDPPIEEHQLR 2396
Qy 251 ADLEQCWDIPNFFQFCNFSVSIVSV-----SETITKMLKA-IEDLPKODKA----- 297
Db 2397 NSIE-----KLARLEADEKKQLCVLQQLKSEHSHADLLKGRVENERLEIARTN 2447
Qy 298 PDHGGL-----ISKMLPGQDRGLCG-ELD-----ONLSRCF----- 327
Db 2448 OEHAALAEANSKGEVETLKAKIEGTMQSLRGLDVTIRSEKENLTNELQKEQERISEL 2507
Qy 328 -----KFHEKCO-----KCOAHLSEDCPDYPALHTE-----L 354
Db 2508 EIINSEFENILQKEQEKVQMKESSTAMEMLOTQLKELNERNVAALHNDQEAKEQNL 2567
Qy 355 DEAIRLVNVSNOQYQ-----ILQMTKHL-----EDTAYLVKMKRGQFGWVSELA 400
Db 2568 SSOVECLELEKAQILQGLDEAKNNYIVLQSSVKGLQOEVEDGKQLEKKDEE---ISRLK 2624
Qy 401 NOAPETEIIFNSIQVVPRIH-----EGNISKQDETMTD--LSILPSSNFTLKIPLE 450
Db 2625 NQIQDQEQVLVSKLSQVEGEHQLWKQONLELRNLTVLEQKIOVLQSKNASLQDTLE 2680

Search completed: July 2, 2002, 11:52:33
Job time: 216 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:52:00 ; Search time 33.04 Seconds

(without alignments)
1387.246 Million cell updates/sec

Title: US-09-722-544A-4MOD

Perfect score: 2529

Sequence: 1 MRTWDYNSGNMKRPLLVFL.....FIGYVAKALQHFKEHFXTW 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length DB	ID	Description
1	434	17.2	449	1 A41386	clusterin precursor
2	418	16.5	451	2 I50131	clusterin - quail
3	410.5	16.2	445	2 A40018	clusterin precursor
4	408	16.1	449	2 S07714	T64 protein precursor
5	403	15.9	448	2 A40714	clusterin precursor
6	402	15.9	448	2 I56335	apolipoprotein J -
7	400.5	15.8	439	2 A35744	clusterin precursor
8	400.5	15.8	446	2 A42108	clusterin precursor
9	380	15.0	447	2 A27205	clusterin precursor
10	188	6.6	191	2 T48174	sulfated glycoprot
11	154	6.5	1738	2 T14867	interaplin - slime
12	151	6.0	1526	2 T41522	myosin II - fission
13	150	5.9	1156	2 B70356	chromosome assembl
14	149	5.9	1388	2 S70633	serine/threonine-s
15	148.5	5.9	1091	2 T34107	hypothetical prote
16	147	5.8	3225	2 I52300	giantin - human
17	147	5.8	3259	1 A56539	giantin - human
18	145.5	5.6	1004	1 A44073	C1K1 protein - yea
19	141.5	5.6	1005	2 A64465	hypothetical prote
20	141.5	5.6	3433	1 S28381	utrophin - human
21	140.5	5.6	2166	2 G70163	hypothetical prote
22	140	5.5	1132	2 T00259	hypothetical prote
23	139.5	5.5	2863	1 S28261	centromere protein
24	138	5.5	1084	2 G71329	hypothetical prote
25	137	5.5	1955	2 T30934	myosin heavy chain
26	137	5.4	1940	1 S04090	myosin heavy chain
27	136	5.4	1388	2 S74245	serine/threonine-s
28	135.5	5.4	1679	2 S48385	hypothetical prote
29	135.5	5.4	1956	2 T16416	hypothetical prote

30	135	5.3	1060	1 A40264	kinesin-related pr
31	135	5.3	1133	2 T22976	hypothetical prote
32	134.5	5.3	1300	2 I53799	Cg1 protein - huma
33	134	5.3	1538	2 T28095	cardiac muscle fac
34	133.5	5.3	1313	2 A48467	myosin heavy chain
35	133.5	5.3	1410	1 A57013	early endosome ant
36	133.5	5.3	1957	2 T38077	hypothetical colle
37	132.5	5.2	852	2 D72230	conserved hypothet
38	132	5.2	1033	2 T42701	hypothetical prote
39	132	5.2	1992	2 A47297	myosin heavy chain
40	131.5	5.2	1979	1 S03166	myosin heavy chain
41	131	5.2	1024	2 T34517	kinesin-related pr
42	131	5.2	1413	2 T26467	hypothetical prote
43	130.5	5.2	1940	1 A24922	myosin heavy chain
44	130	5.1	746	2 T21277	hypothetical prote
45	130	5.1	793	2 S34830	kinesin-related pr

ALIGNMENTS

RESULT 1
A41386
clusterin precursor [validated] - human
N:Alternate names: apolipoprotein J; complement cytolysis inhibitor SP-40; complement TRPM-2/clusterin protein
N:Contains: clusterin alpha chain; clusterin beta chain
C:Species: Homo sapiens (man)
C>Date: 03-Apr-1992 #sequence_revision 17-Nov-1995 #text_change 08-Dec-2000
C/Accession: S43646; S04652; A41386; A35833; S34056; A53177; A37816; B37816; PL0136;
R:Wong, P.; Taillefer, D.; Laking, J.; Pineault, J.; Chader, G.; Tenniswood, M.
Eur. J. Biochem. 221, 917-925, 1994
A>Title: Molecular characterization of human TRPM-2/clusterin, a gene associated with
A:Reference number: S43646; MUID:94237156
A:Accession: S43646
A:Molecule type: DNA
A:Residues: 1-449 <WON>
A:Cross-references: GB:M64722; NID:9339972; PIDN:AAB06508.1; PID:9339973
R:Kirschbaum, L.; Sharpe, J.A.; Murphy, B.; d'Aplice, A.J.F.; Classon, B.; Hudson, P.;
EMBO J. 8, 711-718, 1989
A>Title: Molecular cloning and characterization of the novel, human complement-associ
A:Reference number: S04652; MUID:89251601
A:Accession: S04652
A:Molecule type: mRNA
A:Residues: 1-449 <KIR>
A:Cross-references: EMBL:X14723; NID:930250; PIDN:CAA32847.1; PID:930251
A>Note: parts of this sequence, including the amino end of the mature protein, were c
R:Jenne, D.E.; Tschopp, J.
Proc. Natl. Acad. Sci. U.S.A. 86, 7123-7127, 1989
A>Title: Molecular structure and functional characterization of a human complement cy
lis fluid.
A:Reference number: A41386; MUID:89386692
A:Accession: A41386
A:Molecule type: mRNA
A:Residues: 2-449 <JEN>
A:Cross-references: GB:M25915; NID:9180619; PIDN:AAA51765.1; PID:9180620
R:de Silva, H.V.; Harmony, J.A.K.; Stuart, W.D.; Gil, C.M.; Robbins, J.
Biochemistry 29, 5380-5389, 1990
A>Title: Apolipoprotein J: structure and tissue distribution.
A:Reference number: A35833; MUID:90344779
A:Accession: A35833
A:Molecule type: mRNA
A:Residues: 34-449 <DES>
A:Cross-references: GB:J02908; NID:q178854; PIDN:AAA51765.1; PID:q178855
R:Ghis, J.; Matsubara, E.; Koudinov, A.; Choi-Mitura, N.H.; Tomlita, M.; Wisniewski, T
Biochem. J. 293, 27-30, 1993
A>Title: Research Communication. The cerebrospinal-fluid soluble form of Alzheimer's
A:Reference number: S34056; MUID:93319521
A:Accession: S34056
A:Molecule type: protein
A:Residues: 328-240, 'X', 242-246; 23-24, 'X', 26-34, 'X', 36-38, 'X', 40-41 <GH>
R:James, R.W.; Hochstrasser, A.C.; Borghini, I.; Martin, B.; Pometta, D.; Hochstrasse
Arterioscler. Thromb. 11, 645-652, 1991

A:Title: Characterization of a human high density lipoprotein-associated protein, NA1/NA
A:Reference number: A53177; MUID:91230083
A:Accession: A53177
A:Molecule type: protein
A:Residues: 229-242:303-304, 'M', 306-312, 'X', 314-317,397-403 <IAM>
R:de Silva, H.V.; Stuart, M.D.; Park, Y.B.; Mao, S.J.T.; Gill, C.M.; Wetterau, J.R.; Busc
J. Biol. Chem. 265, 14292-14297, 1990
A:Title: Purification and characterization of apolipoprotein J.
A:Reference number: A37816; MUID:90354412
A:Accession: A37816
A:Molecule type: protein
A:Residues: 23-46, 'H', 48-51, 'Q', <DE3>
A>Note: amino end of the alpha chain
A:Accession: B37816
A:Molecule type: protein
A:Residues: 220-257 <DE2>
A>Note: amino end of the beta chain
R:Choi, N.H.; Mazda, T.; Tomita, M.
Mol. Immunol. 26, 835-840, 1989
A:Title: A serum protein SP40,40 modulates the formation of membrane attack complex of c
A:Reference number: PLO135; MUID:90097955
A:Accession: PLO136
A:Molecule type: protein
A:Residues: 23-37 <CHC>
A>Note: this fragment was isolated from the membrane attack complex SC5b-9
A:Accession: PLO135
A:Molecule type: protein
A:Residues: 228-242 <CH2>
A>Note: this fragment was isolated from the membrane attack complex SC5b-9
R:Hochstrasser, A.C.; James, R.W.; Mattlin, B.M.; Hochstrasser, D.; Pomet
Appl. Theor. Electrophor. 1, 73-76, 1988
A:Title: HDL particle associated proteins in plasma and cerebrospinal fluid: identificat
A:Reference number: S07433; MUID:91265608
A:Accession: S09339
A:Molecule type: protein
A:Residues: 229-240 <HOC>
A:Accession: S07433
A:Molecule type: protein
A:Residues: 24-27, 'S', 29-33 <HOC>
R:Kirsbaum, L.; Bozas, S.E.; Walker, I.D.
FEBS Lett. 297, 70-76, 1992
A:Title: SP-40,40, a protein involved in the control of the complement pathway, possess
A:Reference number: A56293; MUID:92201397
A:Accession: A56293
A:Contents: annotation; disulfide bonds; carbohydrate binding sites
R:Kunitake, S.T.; Carilli, C.T.; Lau, K.; Procter, A.A.; Naya-Vigne, J.; Kane, J.P.
Biochemistry 33, 1988-1993, 1994
A:Title: Identification of proteins associated with apolipoprotein A-I-containing lipop
A:Reference number: A54223; MUID:94162201
A:Accession: E54223
A:Molecule type: protein
A:Residues: 228-246 <KUN>
A:Experimental source: apolipoprotein A-I-containing lipoproteins, plasma
A:Accession: F54223
A:Molecule type: protein
A:Residues: 23-34, 'X', 36-37 <KU2>
A:Experimental source: apolipoprotein A-I-containing lipoproteins, plasma
R:Danik, M.; Chabot, J.G.; Mercier, C.; Benabid, A.L.; Chauvin, C.; Quirion, R.; Sub, M.
Proc. Natl. Acad. Sci. U.S.A. 88, 8577-8581, 1991
A:Title: Human gliomas and epileptic foci express high levels of a mRNA related to rat
A:Reference number: I59206; MUID:92020896
A:Accession: I59206
A>Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 61-449 <RES>
A:Cross-references: GB:M74816; NID:g338056; PIDN:AAA60321.1; PID:g338057
R:Duguid, J.R.; Bohmont, C.W.; Liu, N.G.; Tourtelotte, W.W.
Proc. Natl. Acad. Sci. U.S.A. 86, 7260-7264, 1989
A:Title: Changes in brain gene expression shared by scrapie and Alzheimer disease.
A:Reference number: I48174; MUID:89386721
A:Accession: I63132
A>Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 103-168 <RE2>

A:Cross-references: GB:M26639; NID:g338070; PIDN:AAA36609.1; PID:g553644
C:Comment: This protein has been implicated in complement cascade inhibition, membran
11 as in normal brain, in tissues affected by neurodegenerative disease processes, an
C:Comment: This protein may assist in preventing the formation of Alzheimer's disease
C:Genetics:
A:Gene: GDB:CLU; CLU
A:Cross-references: GDB:125226; OMIM:185430
A:Map position: 8p21-8p21
A>Note: appears to be a single-copy gene; alternative exon usage in 5'-untranslated r
C:Superfamily: clusterin
C:Keywords: apoptosis; complement inhibitor; extracellular protein; glycoprotein; HDL
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-227/Domain: clusterin beta chain #status experimental <BCH>
F:23-227,228-449/Product: clusterin #status experimental <MAN>
F:228-449/Domain: clusterin alpha chain #status experimental <ACH>
F:86,103,145,291,354,374/Binding site: carbohydrate (asn) (covalent) #status experime
F:102-313,121-295,129-285/Disulfide bonds: #status experimental
F:113-305,116-302/Disulfide bonds: (or 113-302, 116-305) #status experimental
F:317/Binding site: carbohydrate (asn) (covalent) #status absent

Query Match 17.28; Score 434; DB 1; Length 449;
Best Local Similarity 25.88; Pred. No. 4.36-20;
Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

QY 12 MKPPLVFIYVCLMLKSDHCAPTWK-----DKTAISNLKSFSGVIGIDAEYKKAAT 65
DB 1 MKKTLILFVGLL-----TWESGYLDQYVSDNEIDEMSGKYNNKELQNNVN 51
QY 66 GIKQKIMMERKEKHTLMSTLKKREKQDAIKLNEVOHELEERELCRESLASWG 125
DB 52 GVKQIKTLIEKTNBERKTLNLNEBAKKKEDALNETRESFKLGLPGVGNETFMALWE 111
QY 126 ECRSLENNCKRIT-TCQPMSSYKKNIEKFFKITYFLPPHEDNEKDLPISEKLEK 184
DB 112 ECKPLKQTCMKFYARVCRSSGLVGRQLEELWQSSPFYMMNGDR-----IDSLLEN 165
QY 185 DAOLITMEDV----FSQITLVNLSLFNRSFNVFMQMOEFOQTQSPHSIS-DTDLTEPYF 239
DB 166 DRQTHMDLVQMDHFSRKSSTIDELFQDRF-----FTRPQDIT--HYLPSPSLHRRPHF 218
QY 240 FPAFSKEMPTADLE-QCWDIPNFPOLFNCNFSVSIYESVETTKMLKAIEDLPKQKAP 298
DB 219 F--FPKSRIVRSILMFPFSYEPLEFNHAFQPLEMHEA-----QAMDIHFHSPAF 267
QY 299 DHGGLISMLRPGDGLGELDONLSRCFHEKCOQAHLSBDC-----PDVPAHTEL 354
DB 268 QHPTEPIREDDDRYVCREIRHNSITGCLRNKDKDCREILSVDCSTNNSQAKLRREL 327
QY 355 DEAIRLVNVSNOQYGOILQMTKRLHEPTAVLYEKMRGQFGVSELANQAPETEIFNSIQ 414
DB 328 DESIQVABERLRRKYNELKSYQMKMLNTSSLLDQNLQDFNVSLANLJGQEDDYILRVY 387
QY 415 VVPRTHGCSNKQDPTMTDILSPSSNFYTKIPLESASSNFYGVAKALQHF-KEH 473
DB 388 TVAS-HTSDSDPVSGVTEVVVKLEPDSPTITVTVPEVRSRKNPKMETVAEKALQDEYRKH 446
RESULT 2
Clusterin - quail
C:Species: Coturnix coturnix (quail)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999
C:Accession: I50131
R:Michel, D.; Chatalein, G.; Hernalt, Y.; Brun, G.
Eur. J. Biochem. 229, 215-223, 1995
A:Title: The expression of the avian clusterin gene can be driven by two alternative
A:Reference number: I50131; MUID:95262670
A:Accession: I50131
A>Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-451 <MIC>
A:Cross-references: EMBL:X80760; NID:g520629; PIDN:CAA56733.1; PID:g520630

C:Genetics:
A:Gene: T64
A:introns: 29/1; 78/3; 135/3; 275/1; 313/1; 388/3; 450/2
C:Superfamily: clusterin

Query Match 16.5%; Score 418; DB 2; Length 451;

Best Local Similarity 25.7%; Pred. No. 4.5e-19;
Matches 118; Conservative 94; Mismatches 179; Indels 68; Gaps 15;

```

QY 45 LKSFSEVGEIDADEYKKAITGKIMMERKEKHTNIMSLTKCREKQKALALNE 104
DB 27 LKQLSAGSKYIDAEVENALNGVKOMKTLMDSKSEHQAHLTEETKKKEAVKLALE 86
QY 105 VOEHLEERLCRESLADSWGECRSCLENNCMRIY-TCOPSSVYKNKIERFFKRYOF 163
DB 87 KEKQLAEKQEVCEHTLSLMEBCKPCLKHTCMYVSKMCHSGGLVGRQLEELNNSSPF 146
QY 164 LPPFHEDNEKDLPISEKLEK-----DAQLTQMEDVFSQLTVDVNSLFNRSFNVRQMOQ 218
DB 147 SIWNGERIDDLDRQGRERFEDLEERFGLMEDGVEDIFQDSTQLYGPAFFPFR---- 202
QY 219 EFDQTFQSHFISDTDLTEPY--FFPAFSKPEPMTKADLEQCMIDPN-----FFQLRC 267
DB 203 -----TPPEGGFREAF-VPPVQKVAH-----VPRRLSRELHPEFQHPM 240
QY 268 NESVSIVSEYSETITKMLKAIEDLPKQDKAPDH--GGLISKMLP--GQDRGLGCLDQNL 324
DB 241 HGHNRFLFQPLFEMTQMLDG-----GHGMEHPLGFPATESNFTDRMYCEIRNNSA 294
QY 325 RCFKFEKQCKQCAHLSEDC---PVPALHTELDPAIRLVNSNOQYGILOMTKHKLE 380
DB 295 GCLRMDECEKCEKREILAVDCSQTPVQSOQLREQFEDALRIAEFRTHRYDILSAFAQEM 354
QY 381 DPAYLEKMRGQGWVSELANQAP-----ETELIFNSIQVVRHIEGNISKODETMMMD 434
DB 355 NTSLLDQDLARQFGWVSRIGNLTGQNDGLQVTTVESK--TPNLEDP--SAPADPQVLT 408
QY 435 LSTLPSSNFTLKIPLEESAESSNFICYVAKALQHFKEH 473
DB 409 VOLFDSEPLSLTYPGDISMDPFRMEIIVAEQAQHYKQN 447

```

RESULT 3
A40018
clusterin precursor - dog
N:Alternate names: 80k secreted glycoprotein, renal; complement cytotoxic inhibitor
C:Species: Canis lupus familiaris (dog)
C>Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 13-Aug-1999
C:Accession: A40018
R:Hartmann, K.; Rauch, J.; Urban, J.; Parczyk, K.; Diel, P.; Pillarsky, C.; Appel, D.; Ha
J. Biol. Chem. 266, 9924-9931, 1991
A:Title: Molecular cloning of gp 80, a glycoprotein complex secreted by kidney cells in
A:Reference number: A40018; MUID:91236776
A:Accession: A40018
A:Molecule type: mRNA
A:Residues: 1-445 <HAR>
A:Cross-References: GB:M55251; GB:M38757; NID:g163953; PIDN:AAA30846.1; PID:g163954
A:Note: parts of this sequence, including the amino ends of the alpha and beta chains, w
C:Superfamily: clusterin
C:Keywords: apoptosis; disulfide bond; glycoprotein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-226/Domain: clusterin beta chain #status experimental <BCH>
F:23-226,227-445/Product: clusterin #status experimental <MAT>
F:227-445/Domain: clusterin alpha chain #status experimental <ACH>
F:86,103,145,277,287,350,370/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 16.2%; Score 410.5; DB 2; Length 445;
Best Local Similarity 23.9%; Pred. No. 1.3e-18;
Matches 117; Conservative 102; Mismatches 192; Indels 79; Gaps 15;

QY 12 MKPPLLVIVCLIMLADSHCAPTWKDKTAISENLKSFSEVGEIDAD-----EEVKKALT 65

```

DB 1 MKKTLILLVGLL-----TWNGRVLGQAVSDVELQEMSTEGSKYINKEIKNAK 51
QY 66 GIKOKIMMERKEKHTNIMSLTKCREKQKALALNEVOEHLEERLCRESLADSWG 125
DB 52 GKQKLTLEQNEERKSLSLNLEEKKKEDALNDTKSEFKLASQGVCDTMAALME 111
QY 126 ECRSCLENNCMRIY-TCOPSSVYKNKIERFFKRYOFLPPFHEDNEKDLPISEKLEK 184
DB 112 ECKPCLQKQCMFYARVCSGGLVGHQLEEFNLNOSPPEYFMWNGDR-----IDSLLEN 165
QY 185 DAQLTQMEDVFSQLTVDVNSLFNRSFNVRQMOQEPDQTFQSHFIS-----DTDLTEPYF 240
DB 166 DRQQTALDV-----MODSFRASSI-----MDELFORFFRQPDYHYSP--F 209
QY 241 PAFSKPEPMTKADLEQCMID-----PNEFOLCFNFSVISEYSETITKMLKAIE--- 289
DB 210 SLFORRPFNPKFRIARNIIPPRFQPLNFHDQFFQPMIQA-QQANDVLRHPIYHF 268
QY 290 --DLPKQDKAPDHGGLISKMLPQDRGLGCLDQNLsrcfKFEKQCKQCAHLSEDC--- 344
DB 269 PLIEPRED-----NRTVCKEIRHNSGTGLKMKDQCEKCOELLISVDCSSN 312
QY 345 -PVPALHTELDPAIRLVNSNOQYGILOMTKHKLEDPALYVEKMRGQGWVSELANQ 403
DB 313 NPAQVLRQELNSLQIAEKFTKLYDELQSYQEKMFNTSSLKQLEQFSWVSQLANLT 372
QY 404 PETELIFNSIQV-PRHIEGNISKODETMMTSLPSSNFTLKIPLEESAESSNFICYV 462
DB 373 QSEDPFYLOVTTVYGSTQSDSNVPVGFYKYV--VKLSDSPITVIMPEAVSRNNPFMEY 430
QY 463 VAKALQHFKE 472
DB 431 AEKALQGYRQ 440

```

RESULT 4
S07714
T64 protein precursor - Japanese quail
C:Species: Coturnix coturnix japonica (Japanese quail)
C>Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 13-Aug-1999
C:Accession: S07714
R:Michel, D.; Gillet, G.; Volovitch, M.; Pessac, B.; Calothy, G.; Bruun, G.
Oncogene Res. 4, 127-136, 1989
A:Title: Expression of a novel gene encoding a 51.5 kD precursor protein is induced b
A:Reference number: S07714; MUID:89239492
A:Accession: S07714
A:Molecule type: mRNA
A:Residues: 1-449 <MIC>
A:Cross-References: EMBL:X15825; NID:g62594; PIDN:CAA33823.1; PID:g62595
C:Superfamily: clusterin
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-449/Product: T64 protein #status predicted <MAT>

Query Match 16.1%; Score 408; DB 2; Length 449;
Best Local Similarity 25.7%; Pred. No. 1.9e-18;
Matches 117; Conservative 92; Mismatches 179; Indels 68; Gaps 15;

```

QY 45 LKSFSEVGEIDADEYKKAITGKIMMERKEKHTNIMSLTKCREKQKALALNE 104
DB 27 LKQLSAGSKYIDAEVENALNGVKOMKTLMDSKSEHQAHLTEETKKKEAVKLALE 86
QY 105 VOEHLEERLCRESLADSWGECRSCLENNCMRIY-TCOPSSVYKNKIERFFKRYOF 163
DB 87 KEKQLAEKQEVCEHTLSLMEBCKPCLKHTCMYVSKMCHSGGLVGRQLEELNNSSPF 146
QY 164 LPPFHEDNEKDLPISEKLEK-----DAQLTQMEDVFSQLTVDVNSLFNRSFNVRQMOQ 218
DB 147 SIWNGERIDDLDRQGRERFEDLEERFGLMEDGVEDIFQDSTQLYGPAFFPFR---- 202
QY 219 EFDQTFQSHFISDTDLTEPY--FFPAFSKPEPMTKADLEQCMIDPN-----FFQLRC 267

```

Db 203 -----TPPFGCFREAF-VPRQVRHL-----VPRRLSREHLHPFHQM 240

Qy 268 NESVATESVSTITKMLKAIEDLPKODKAPDH--GGLISKMLP-GODGCLGELDONS 324

Db 241 HGFHRLFOPLFMTQMLDGC-----GHGAMHPLGFGFATESRNMSTDMVRELIRNSA 294

Qy 325 REFKEHKCQKQAHLEDC-----PDVPLALTELEDAIRLVNVSQOQIQTMRKILE 380

Db 295 GCLRNRDECEKRELAVADCSQTDVQSQLRQFEDALRLERFTRRYRDDLLSAFQAEML 354

Qy 381 DTAVYAEKRGQGVNSELANOAP-----ETELFENSIQVVPRIHEGNISKODEMTMD 434

Db 355 NNSSLDQNRFGVNSRGLNLTQGNDFLOVTTVFSK---TPNLEDP--SAPADTOYT- 408

Qy 435 LSLPSSNFTKIPLEESAESSNFTGYVAKALQHF 470

Db 409 VOLFDSEPLSLTVPGDISWMDPRFMEIWEDEALQHY 444

RESULT 5

clusterin precursor - mouse
A:Accession: A40714
N:Alternate names: sulfated glycoprotein 2
C:Contains: clusterin alpha chain; clusterin beta chain
C:Species: Mus musculus (house mouse)
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 13-Aug-1999
R:Accession: A40714; JN0699
R:RefSeq: L.E.; Chomn, A.; Duceast, D.; Baumann, B.; Bellin, D.; Wohlwend, A.; Kiss, J.Z.
J. Cell Biol. 122, 1119-1130, 1993
A:Title: Murine clusterin: molecular cloning and mRNA localization of a gene associated
A:Reference number: A40714; MUID:93355508
A:Accession: A40714
A:Molecule type: mRNA
A:Residues: 1-448 <FRR>
A:Cross-references: GB:I08235; NID:g192596; PIDN:AAA37422.1; PID:g192597
R:Lee, K.H.; Ji, Y.M.; Lam, H.M.; Lee, S.C.; You, K.H.
Biochem. Biophys. Res. Commun. 194, 1175-1180, 1993
A:Title: Molecular cloning and sequencing of sulfated glycoprotein-2 cDNA from testis of
A:Reference number: JN0699; MUID:93356785
A:Accession: JN0699
A:Molecule type: mRNA
A:Residues: 1-448 <LEE>
A:Note: the authors translated the codon CGC for residue 66 as Ala
C:Comment: This protein is involved in many different cellular processes other than those
C:Superfamily: clusterin
C:Keywords: apoptosis; disulfide bond; glycoprotein; spermatogenesis
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-226/Domain: clusterin beta chain #status predicted <RCH>
F:227-226,227-448/Product: clusterin #status predicted <MAY>
F:227-448/Domain: clusterin alpha chain #status predicted <ACH>
F:102,144,290,327,353,373/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 15.9%; Score 403; DB 2; Length 448;
Best Local Similarity 23.4%; Pred. No. 3,9e-18;

Matches 112; Conservative 109; Mismatches 196; Indels 62; Gaps 14;

Qy 16 LLLVIVCLMLKDSHCAPTMDKTAISENLKSFSEVGEIDA-----DEYKALTLGKIQ 69

Db 3 ILLCVALLLI-----WDNGMVLGEQVSDNELQELSTQSGRYINKETIONAVQYKH 54

Qy 70 MKIMMERKEKHTNLMSTLKCRREKQKALKLNEVOEHLSEERLCRESLADSGECRS 129

Db 55 IKTLKRNKRLSLNSLEAKKKEDALDTRDSEKMLKAPREVCNEMTMALMECKRP 114

Qy 130 CLENNCMRIYT-TCOPSSSVYKNIERFRKTYOFLPFHEDNEKDLPISEKLEKDAQ 188

Db 115 CLKHTCKMFYARVCRSGGLVGOQLEEFLLNGSPFFYFMNDR-----IDSLLESRDQ 168

Qy 189 TO-----MEDVSQTLVDVNSLFRNSFNVRQMOQEFDTQSHFISDTLPEYFP-- 241

Db 169 SQVLDAMQDSFARASGIDTLFDORF--FAR-----ELHDPHYFS-----PIGFPHKR 214

Qy 242 ---AFSEKPMRKADLEQC-WDIPNFFOLCFNFSVATESSTITKMLKAIEDLPKODKA 297

Db 215 PHFLPKSLNVLNLSPSHYGPPSPFNMFQPFEEKHQAQAQMDVQLSPAFQFDVD-- 272

Qy 298 PDHGLISKMLPGQ-DKGLCGELDONLSRCRFRHEKCKQAHLSDC-----PDVPLALHT 352

Db 273 -----FLREGEDRNVCKEIRRNSTGCLMKQCEKQCEILISVDCSTNNPAQANRQ 324

Qy 353 ELDEAIRLVNVSNOOQIOLMTKRLHEDTAYLVKMGQGVNSELANOAPETELFINS 412

Db 325 ELNDSLOVAERLTQYKELQSFQSKMLNTSLLEQLNDQFNWVSQALNLTQGEDEKTYLR 384

Qy 413 IQVPRIRHEGNISKODEMTMDLSILPSSNFTKIPLEESAESSNFTGYVAKALQHF 471

Db 385 VSTV-TTHSSDSEVPSRYTEVYVVKLFDSDPITVYLPREVSKDNPRFMDTVAEKALQETR 442

RESULT 6

apolipoprotein J - mouse
A:Accession: 156335
C:Species: Mus sp. (mouse)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 13-Aug-1999
R:Jordan-Scarck, T.C.; Lund, S.D.; Witte, D.P.; Aronow, B.J.; Ley, C.A.; Stuart, W.D.
J. Lipid Res. 35, 194-210, 1994
A:Title: Mouse apolipoprotein J: characterization of a gene implicated in atherosclerosis
A:Reference number: 156335; MUID:94222204
A:Accession: 156335
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-448 <RES>
A:Cross-references: GB:S70244; NID:g546504; PIDN:AAB30623.1; PID:g546505
A:Genetics:
A:Gene: ApolJ
C:Superfamily: clusterin

Query Match 15.9%; Score 402; DB 2; Length 448;
Best Local Similarity 24.1%; Pred. No. 4,5e-18;

Matches 115; Conservative 108; Mismatches 201; Indels 54; Gaps 15;

Qy 12 MKPPLVIVCLMLKDSHCAPTMDKTAISEN-LKSESEVGEIDADEVKALTLGKIQ 70

Db 1 MKILLCVAMLLIW--DNGMVLGEQVSDNELQELSTQSGRYINKETIONAVQYKH 55

Qy 71 KIMMERKEKHTNLMSTLKCRREKQKALKLNEVOEHLSEERLCRESLADSGECRS 130

Db 56 KTLKRNKRLSLNSLEAKKKEDALDTRDSEKMLKAPREVCNEMTMALMECKRP 115

Qy 131 LENNCMRIYT-TCOPSSSVYKNIERFRKTYOFLPFHEDNEKDLPISEKLEKDAQ 189

Db 116 LKHTCKMFYARVCRSGGLVGOQLEEFLLNGSPFFYFMNDR-----IDSLLESRDQ 169

Qy 190 Q-----MEDVSQTLVDVNSLFRNSFNVRQMOQEFDTQSHFISDTLPEYFP-- 241

Db 170 QVLDAMQDSFARASGIDTLFDORF--FAR-----ELHDPHYFS-----PIGFPHKR 215

Qy 242 ---AFSEKPMRKADLEQC-WDIPNFFOLCFNFSVATESSTITKMLKAIEDLPKODKA 298

Db 216 HFLPKSLNVLNLSPSHYGPPSPFNMFQPFEEKHQAQAQMDVQLSPAFQFDVD-- 272

Qy 299 DHGLISKMLPGQ-DKGLCGELDONLSRCRFRHEKCKQAHLSDC-----PDVPLALHT 353

Db 273 -----FLREGEDRNVCKEIRRNSTGCLMKQCEKQCEILISVDCSTNNPAQANRQ 325

Qy 354 LDEAIRLVNVSNOOQIOLMTKRLHEDTAYLVKMGQGVNSELANOAPETELFINS 413

Db 326 LNDLSLOVAERLTQYKELQSFQSKMLNTSLLEQLNDQFNWVSQALNLTQGEDEKTYLR 385

Qy 414 QVPRIRHEGNISKODEMTMDLSILPSSNFTKIPLEESAESSNFTGYVAKALQHF 471

Db 386 STV-TTHSSDSEVPSRYTEVYVVKLFDSDPITVYLPREVSKDNPRFMDTVAEKALQETR 442

A:Accession: A45890; S18491; A45415; B31575; A31575; A27205
R:Collard, M.W.; Griswold, M.D.
Biochemistry 26, 3297-3303, 1987
A:Title: Biosynthesis and molecular cloning of sulfated glycoprotein 2 secreted by rat
A:Reference number: A45890; MUID:88000523
A:Accession: A45890
A:Molecule type: mRNA
A:Residues: 1-447 <CO2>
A:Cross-references: GB:M16975; NID:g204472; PIDN:AA441273.1; PID:g204473
R:Betluzzi, S.; Hlupka, R.A.; Gilina, P.; Liao, S.
Biochem. J. 257, 293-296, 1989
A:Title: Identification of an androgen-repressed mRNA in rat ventral prostate as coding
A:Reference number: S18491; MUID:89149740
A:Accession: S18491
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-186, 'D', 188-447 <BET>
A:Cross-references: EMBL:X13231; NID:g57240; PIDN:CA31618.1; PID:g57241
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988
J. Mol. Biol. 208, 5021-5031, 1993
U. Biol. Chem. 268, 5021-5031, 1993
A:Title: Genomic organization and expression of the rat TRPM-2 (clusterin) gene, a gene
A:Reference number: A45415; MUID:93186813
A:Accession: A45415
A:Status: preliminary
A:Molecule type: DNA; mRNA
A:Residues: 1-447 <WON>
A:Experimental source: blood
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBI:126803, NCBI:126805)
R:Cheng, C.Y.; Chen, C.L.C.; Feng, Z.M.; Marshall, A.; Bardini, C.W.
Biochem. Biophys. Res. Commun. 155, 398-404, 1988
A:Title: Rat clusterin isolated from primary Sertoli cell-enriched culture medium is su
A:Reference number: A90146; MUID:86326333
A:Accession: B31575
A:Molecule type: protein
A:Residues: 22-51 <CHE>
A:Note: amino end of the beta chain
A:Accession: A31575
A:Molecule type: protein
A:Residues: 227-256 <CH2>
A:Note: amino end of the alpha chain
C:Superfamily: clusterin
C:Keywords: Apoptosis; disulfide bond; glycoprotein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-226/Domain: clusterin beta chain #status experimental <BCH>
F:227-226, 227-447/Product: clusterin #status experimental <MAT>
F:227-447/Domain: clusterin alpha chain #status experimental <ACH>
F:102, 144, 290, 327, 353, 373/Binding site: carbohydrate (Asn) (covalent) #status predicted

Db	215	PHFYPSRLVRSILMPESHVGRPLSEFNNMFOEFPMIHQAQAMDOVUHLSPALQEPDDEL	274
QY	298	PDHGILSKMLPGQ-DKGLGELDQNTSRCEKHEKQCAKCAHLSEDC----	PDVALAHT 3523
Db	275	KE-----GDDPTVCKEIRHNSTGCKKMGQCEKCEILSYDCSTNNPAQANLRQ	324
QY	353	ELDEAIRLVNVSNOQYCOILQMTKRHLIEDPAVLVEKKRGQFGWSELAN--QAPEEIIIF	410
Db	325	ELNLSQVAEKLTQCYVELLHSLQSKMLNTSSLLEQLNDQFTWVSQLANLTQGGDOYLRY	384
QY	411	NSIOVPRIEHGNTSKQDETMMDLSILPSSNFTKLIPEESASSNFIQYVAKALQHF	470
Db	385	STVTT----HSDSEVPSRVTEVVVKLFSDDPITVVLPEEVSXNDNPKMDTVAKALQEX	440
QY	471	K 471	
Db	441	R 441	

RESULT	10
148174	sululated glycoprotein 2 - golden hamster (fragment)
C:Species:	Mesocricetus auratus (golden hamster)
C:Date:	27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 13-Aug-1999
C:Accession:	I48174
C:Duguid, J.R.; Bohmont, C.W.; Liu, N.G.; Tourtellotte, W.W. Proc. Natl. Acad. Sci. U.S.A. 86, 7260-7264, 1989	
A:Title:	Changes in brain gene expression shared by scrapie and Alzheimer disease
A:Reference number:	I48174; MUID:89386721
A:Accession:	I48174
A:Status:	preliminary; translated from GB/EMBL/DBJ
A:Molecule type:	mRNA
A:Residues:	1-191 <RES>
A:Cross-references:	GB:I26660; NID:g191450; PIDN:AAA37102.1; PID:g191451
C:Superfamily:	ClusterIn

[illegible]

RESULT 11	
T14867	
Interaptin - slime mold (Dictyostelium discoideum)	
C:Species: Dictyostelium discoideum	
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000	
C:Accession: T14867	
R:Giviero, F.D.; Kuspa, A.; Birkamp, R.; Matzner, M.; Noegel, A.A.	
J. Cell Biol. 142, 735-750, 1998	
A:Title: Interaptin, an actinin-binding protein of the alpha-actinin superfamily in Dictyostelium	
A:Reference number: Z18248; PMID:98365468	
A:Accession: T14867	
A:Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: DNA	
A:Residues: 1-1738 <RIV>	
A:Cross-references: EMBL:AF057019; NID:93549260; PID:93549261; PIDN:MAC34582.1	
C:Genetics:	
A:Gene: abpD	
A:Introns: 173/2; 1680/1	

```

QY      242 ---AFSKEWMTKADLEOCWDIP-NEFOLFNCFSVSATYESVSSETITKMLKAIEDIPKODKA 2977
          : | : : : | : | : : : : : |

```

Query Match	6.58;	Score 164;	DB 2;	Length 1738;
Best Local Similarity	21.18;	Pred. No. 0.025;		

Matches 103; Conservative 87; Mismatches 171; Indels 128; Gaps 21;

```

QY 36 KDKAISENL-----KSEVEGEIDADEVKKALGIMKOMIMMER----- 76
D 667 KDNQITINDQNLKOLSEKDEKLEKLSNOOEOQODERKINNLLLEKKEKCLLERINOOLLEN 726
QY 77 -----KEKEHTNLMSTLKCREEKQOALKLLNEVOHEEER 114
D 727 IDLNSKYQQLLEFENFKLNSKEKE-----NQNLQSKQDEFFNQLND--EKLEKEKQ 779
QY 115 LCRESLADSNQECSCLENNCMKRIYTCQPSWSVKKIEFFFKITYQLEPPHEDNEKD 174
D 780 L--OSIDERNOYK-----QQOLSSNSN-IDOQLOSTIETSELKEQKELN 822
QY 175 LPISEKLEKDAQLOLQMEDVFSQLT-----VPVNSLNFNSFVFMROMOEFDQFOSH 227
D 823 ---DSKLEKEKQLOQLOQOEFDQLEKNQKDHQDQLELEKQ---LKQLOQEHQOLMETN 876
QY 228 FISDTDLTEPFEPFAFSKEPMTKADLE-----QCWDIPNFQOLFQNFVSIVESY 277
D 877 QSIENQNLQOONL--INKENLNEKEQELKLQNLQOIEKIQDFQOEFKQNSINIELV 933
QY 278 SETTKMLKALIED---LPKQDKA---PDHGLISKMLPGQDRGLCGELDONLSKCFEHE 331
D 934 NEKEKELIQLOQDYDQLOKQNSRSDKENDLEK--ENQLSKSIONELNOLIEKNESEDK 991
QY 332 KCQCKQAHLSDECPD---PVALTELEDE---AIRLVNSNOQYGOIL-----Q 373
D 992 EQQLKQOSTINDLEKENQLOQLOQLOQLOQOOSNLSKEDQOLNOLIEKQFQDEKQ 1051
QY 374 MTRKHELDYV---LVEKRRGQFG--WSELANQAPETETIIFNSIQVVPRIHGNISKOD 428
D 1052 LKQOOSIENDLFEKENQIQLOQLOQLOQOOSNLSKED-----QOLNOLIEKNESEDK 1105
QY 429 ETWMTDLSTI 437
D 1106 EQQLKQOSTI 1114

RESULT 12
T41522
myosin 11 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 08-Sep-2000
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, March 1999
A:Reference number: 222000
A:Accession: T41522
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1526 <MOO>
A:Cross-references: EMBL:AL049498; PIDN:CA839901.1; GSPDB:GN00068; SPDB:SPCC645.05C
A:Experimental source: strain 97zh-; cosmid c645
C:Genetics:
A:Gene: SPDB:SPCC645.05C
A:Map position: 3
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:80-743/Domain: myosin motor domain homology <MMO>

```

Query Match 6.0%; Score 151; DB 2; Length 1526;
 Best Local Similarity 20.6%; Pred. No. 0.14;
 Matches 98; Conservative 73; Mismatches 198; Indels 106; Gaps 17;

```

QY 47 SFSEVEIDADEVKKALGIMKOMIMMERKEHTNLMSTLKCREEKQOALKLLNEVO 106
D 895 SFSEVQ--QNEMLQRESASAKOINNEBELLEKTSKVELT-----LSEQNEK 942
QY 107 EHLEBERLCRESLADSNQECSCLENNCMKRIYTCQPSWSVKNKIERFRKIYQFLFP 166
D 943 EKLSLEEK---DLDTKGELSELRENN-----ATVLSKAEE----- 976

```

```

QY 167 FHEDNEKDLPISEKLEKDAQLOLQMEDVFSQLTVDVNS--LFWNSFVFMROQFED--- 221
D 977 -----NEQCKSLQETIVKQDLQDLTKYISDYKTEIQEMRLTNQKME--KSIQOESLSLSE 1031
QY 222 -----QTFQSHFISDTDLTELE-----YFFPAFSKEPMT 249
D 1032 SLKRVKKLEENSTLISDSVILKQOKEELSVLQVQELTINNLEKYNTLEADVQPL 1091
QY 250 KADLEQCGWDIPNFFQOLFQNFVSIVESSETTKMLKALIEDP--KODKAPDHGLISKML 308
D 1092 KLEELSLNDQLOLYOQATKNKELEAKVKECLNNIKSITKLELNKEKQCNLSQASIKYI 1151
QY 309 PGQDRGLCGELDONL-----SRCKFHEKQCKQAHLSDECPDPVALTELEDAIRLVNS 364
D 1152 ELQ-----EIHENLLKYSDSLQENYKKTGQLDL--ESGLQDVTNFOELSKNRDLTFRN 1204
QY 365 NO-----OYGOILOMTRKHELDYVLEKMRGQFGWVSELANQAPETETIIFNSIQV 417
D 1205 HESILRQASASYKREKLSLASENKDLSNKVSSSLKQYNELSPKSKVPELE-----RKITN 1259
QY 418 RIHE--GNISKQDETMMTDLSTLPSSNFTKIPLEESNESSNFTGYVAKALQHEK 471
D 1260 LMHEYSQLGKTFDEKRAKLASRDNBEIL--SLKSELESKRKLEVEYQKYLEEVK 1313

```

```

RESULT 13
B70356
chromosome assembly protein homolog - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Jun-2000
C:Accession: B70356
R:Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MID:98196666
A:Accession: B70356
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1156 <AQP>
A:Cross-references: GB:AE000699; NID:g2983238; PIDN:AA06839.1; PID:g2983243; GB:AE00
A:Experimental source: strain VFS
C:Genetics:
A:Gene: xcpC
C:Superfamily: chromosome segregation protein SMC1

```

Query Match 5.9%; Score 150; DB 2; Length 1156;
 Best Local Similarity 20.3%; Pred. No. 0.12;
 Matches 90; Conservative 88; Mismatches 162; Indels 104; Gaps 17;

```

QY 45 LKSFSEVGEID-----ADEVKKALGIMKOMIMME-----RKEKEHTNLMSTLKK 90
D 163 IEISIGIEYERKERRKALAELEVELEIKIIDLIEISNOLRKLKEKEKLEKFEKQ 222
QY 91 CREKQOALKLLNEVOHELEEBERLCRE--SLADSNQECSCLENNCMKRIYTCQPSWS 148
D 223 IKRE--TEAKITLLEKREKLLKERRIILNLSLRESIEDTTFQIOEKEKL----- 271
QY 149 VKNKIERFFRKIYQLEPPHEDNEK--DLPISEKLI--EKDAQLOLQMEDVFSQLTVDVNS 204
D 272 --NERERLLKEVNEKIMPEKKGFTAEIENAEKSEIKEREKLEKSEKRVKNLEELINN 329
QY 205 LFWNSFVFMRO--MQOEDQTFQSHFISDTDLTEY--FFPAFSKEPMTKADLEQCGWDIP 260
D 330 LLSDKENLREVGTLQLELEK-----LKEEYKSLKEVERKLELEEE----- 373
QY 261 NFFQOLFQCNF--SVSIVESSETTKMLKALIEDLPKODKAPDHGLISKMLPGQDRGLCGEL 319
D 374 --ERLKITTEDEVKLLKEEKEKLEKINSL--NNEKQLELEQKRNKLNKI-----EKI 421
QY 320 DONLSKCFEHEKQCKQAHLSDECPDPVALTELEDAIRLVNSNOQYGOILOMTRKHL 379

```


Search completed: July 2, 2002, 11:52:02
Job time: 225 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:58:14 ; Search time 17.57 Seconds
(without alignments)
1051.180 Million cell updates/sec

Title: US-09-722-544A-4MOD

Perfect score: 2529

Sequence: 1 MRTWYSNSGNMKPPLVFI.....FIGYVAKALQHFREHFTW 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	434	17.2	449	1	CLUS_HUMAN
2	418	16.5	451	1	CLUS_COTJA
3	413.5	16.4	447	1	CLUS_RABIT
4	410.5	16.2	445	1	CLUS_CANFA
5	403	15.9	448	1	CLUS_MOUSE
6	400.5	15.8	439	1	CLUS_BOVIN
7	400.5	15.8	446	1	CLUS_PIG
8	380	15.0	447	1	CLUS_RAT
9	158	6.6	191	1	CLUS_MESAU
10	151	6.0	1526	1	MY82_SCHPO
11	145.5	5.8	594	1	CIK1_YEAST
12	145.5	5.8	1938	1	MYHD_HUMAN
13	141.5	5.6	1005	1	RA50_METJA
14	141.5	5.6	3433	1	UTRO_HUMAN
15	139.5	5.5	2653	1	GENE_HUMAN
16	138	5.5	5430	1	ACR7_HUMAN
17	137	5.4	1940	1	MYH3_HUMAN
18	137	5.4	2230	1	GOG4_HUMAN
19	135.5	5.4	1679	1	YIO9_YEAST
20	135	5.3	1060	1	EG51_XENLA
21	133.5	5.3	1957	1	YD86_SCHPO
22	132.5	5.2	852	1	RA50_THEMA
23	131.5	5.2	1978	1	MYHB_CHICK
24	130.5	5.2	1940	1	MYH3_RAT
25	130	5.1	793	1	KATA_ARATH
26	129.5	5.1	1427	1	REST_HUMAN
27	129.5	5.1	2077	1	TEGU_HSV6U
28	129	5.1	539	1	MY53_HYDAT
29	128	5.1	1939	1	MYH1_HUMAN
30	128	5.1	1744	1	TANA_XENLA
31	128	5.1	1853	1	MY5A_MOUSE
32	128	5.1	2704	1	BPAL_HUMAN
33	127.5	5.0	722	1	MF1_TOBAC

34	127	5.0	1189	1	SCII_CHICK
35	126.5	5.0	971	1	Y228_BORBU
36	126.5	5.0	1875	1	MYH2_YEAST
37	126	5.0	1941	1	MYH2_HUMAN
38	126	5.0	1966	1	MY5B_CAEEL
39	125.5	5.0	978	1	RA5Q_AQAE
40	125.5	5.0	1067	1	EG52_XENLA
41	125.5	5.0	1969	1	MYSA_CAEEL
42	125.5	5.0	3210	1	CENF_HUMAN
43	125	4.9	1790	1	USO1_YEAST
44	125	4.9	1960	1	MYH9_HUMAN
45	125	4.9	2418	1	SPCA_HUMAN

ALIGNMENTS

RESULT	1
CLUS_HUMAN	
ID	CLUS_HUMAN STANDARD; PRT; 449 AA.
AC	P10309; P11380; P11381;
DT	01-JUL-1989 (Rel. 11, Created)
DT	01-JUL-1989 (Rel. 11, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	Clusterin precursor (Complement-associated protein SP-40,40)
DE	(Complement cytolysis inhibitor) (CLI) (NAL and NA2) (Apolipoprotein J) (Apo-J) (TRPM-2).
GN	CLU.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC	TISSUE=Liver;
RX	MEDLINE=89251601; PubMed=2721499;
RA	Kirsbaum L., Sharpe J.A., Murphy B., D'Apice J.F.A., Classon B.,
RA	Hudson P., Walker I.D.;
RT	"Molecular cloning and characterization of the novel, human
RT	complement-associated protein, SP-40,40: a link between the
RT	complement and reproductive systems.";
RL	EMBO J. 8:711-718(1989).
RN	[2]
RP	SEQUENCE OF 2-449 FROM N.A.
RX	MEDLINE=89386692; PubMed=2780565;
RA	Jenne D.E., Tschopp J.;
RT	"Molecular structure and functional characterization of a human
RT	complement cytolysis inhibitor found in blood and seminal plasma:
RT	identity to sulfated glycoprotein 2, a constituent of rat testis
RT	fluid.";
RL	Proc. Natl. Acad. Sci. U.S.A. 86:7123-7127(1989).
RN	[3]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=90344779; PubMed=1974459;
RA	de Silva H.V., Harmony J.A.K., Stuart W.D., Gil C.M., Robbins J.;
RT	"Apolipoprotein J: structure and tissue distribution.";
RL	Biochemistry 29:5380-5389(1990).
RN	[4]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=94237156; PubMed=8181474;
RA	Wong P., Taillefer D., Lakin J., Pineault J., Chader G.,
RT	"Molecular characterization of human TRPM-2/clusterin, a gene
RT	associated with sperm maturation, apoptosis and neurodegeneration.";
RL	Eur. J. Biochem. 221:917-925(1994).
RN	[5]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=93186813; PubMed=7680346;
RA	Pineault J.M., Tenniswood M.;
RT	"Genomic organization and expression of the rat TRPM-2 (clusterin)
RT	gene, a gene implicated in apoptosis.";
RL	J. Biol. Chem. 268:5021-5031(1993).
RN	[6]

RP SEQUENCE FROM N.A.
RA Blechschmidt K., Rosenthal A.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN [7]
RP SEQUENCE OF 61-449 FROM N.A.
RX MEDLINE-92020896; PubMed-1924317;
RA Danik M., Chabot J.G., Mercier C., Benabid A.L., Chauvin C.,
RL Quirion R., Suh M.;
RT "Human gliomas and epileptic foci express high levels of a mRNA
related to rat testicular sulfated glycoprotein 2, a purported marker
of cell death.";
RN Proc. Natl. Acad. Sci. U.S.A. 88:8577-8581(1991).
RN [8]
RP SEQUENCE OF 140-449 FROM N.A.
RC TISSUB=Liver;
RA Glew M.D., Kirszbaum L., Bozas S.E., Walker I.D.;
RL Submitted (JAN-1993) to the EMBL/GenBank/DDBJ databases.
RN [9]
RP SEQUENCE OF 23-33 AND 228-240.
RX MEDLINE-91265608; PubMed-3154963;
RA Hochstrasser A.-C., James R.W., Martin B.M., Harrington M.,
RL Hochstrasser D.F., Pometta D., Merril C.R.;
RT "HDL particle associated proteins in plasma and cerebrospinal fluid:
identification and partial sequencing.";
RN Appl. Theor. Electrophor. 1:73-76(1988).
RN [10]
RP SEQUENCE OF 23-33; 229-242; 303-317 AND 397-403.
RX MEDLINE-91230083; PubMed-1903064;
RA James R.W., Hochstrasser A.-C., Borghini I., Martin B.M., Pometta D.,
RL Hochstrasser D.F.;
RT "Characterization of a human high density lipoprotein-associated
protein, NAI/NA2. Identity with SP-40,40, an inhibitor of complement-
mediated cytotoxicity.";
RN Arterioscler. Thromb. 11:645-652(1991).
RN [11]
RP SEQUENCE OF 23-52 AND 228-257.
RX MEDLINE-90354412; PubMed-2387851;
RA de Silva H., Stuart W.D., Park Y.B., Mao S.J.T., Gil C.M.,
RL Wetterau J.R., Busch S.J., Harmony J.A.K.;
RT "Purification and characterization of apolipoprotein J.";
RN J. Biol. Chem. 265:14292-14297(1990).
RN [12]
RP SEQUENCE OF 23-37 AND 228-242.
RX MEDLINE-90097955; PubMed-2601725;
RA Choi N.H., Mazda T., Tomita M.;
RL "A serum protein SP40,40 modulates the formation of membrane attack
complex of complement on erythrocytes.";
RN Mol. Immunol. 26:835-840(1989).
RN [13]
RP SEQUENCE OF 23-41 AND 228-246.
RX MEDLINE-93319521; PubMed-8328966;
RA Ghiso J., Matsubara E., Koudinov A., Choi-Miura N.-H., Tomita M.,
RL Wisniewski T., Frangione B.;
RT "The cerebrospinal-fluid soluble form of Alzheimer's amyloid beta is
complexed to SP-40,40 (apolipoprotein J), an inhibitor of the
complement membrane-attack complex.";
RN Biochem. J. 293:27-30(1993).
RN [14]
RP PARTIAL SEQUENCE, DISULFIDE BONDS, AND CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE-92201397; PubMed-1551440;
RA Kirszbaum L., Bozas S.E., Walker I.D.;
RL "SP-40,40, a protein involved in the control of the complement
pathway, possesses a unique array of disulphide bridges.";
RN FEBS Lett. 297:70-76(1992).
RN [15]
RP DISULFIDE BONDS.
RX MEDLINE-93147076; PubMed-1491011;
RA Choi-Miura N.H., Takahashi Y., Nakano Y., Tobe T., Tomita M.;
RL "Identification of the disulfide bonds in human plasma protein
SP-40,40 (apolipoprotein-J).";
RN J. Biochem. 112:557-561(1992).
RN [16]
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE-97477902; PubMed-9336835;
RA Kapron J.T., Hilliard G.M., Lakins J.N., Tenniswood M.P., West K.A.,
RL Carr S.A., Crabb J.W.;
RT "Identification and characterization of glycosylation sites in human
serum clusterin.";
RN Protein Sci. 6:2120-2133(1997).
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CELL DEATH (APOPTOSIS).
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X14723; CAA32847.1; -;
CC EMBL; M25915; AAA35692.1; -; ALT_INIT.
CC EMBL; J02908; AAA31765.1; -; ALT_INIT.
CC EMBL; M74816; AAA60321.1; -;
CC EMBL; M64722; AAB06508.1; -;
CC EMBL; L00974; AAA60567.1; -;
CC EMBL; AF311103; -; NOT_ANNOTATED_CDS.
CC EMBL; M63379; AAB06507.1; -;
CC EMBL; M63376; AAB06507.1; JOINED.
CC EMBL; M63377; AAB06507.1; JOINED.
CC EMBL; M63378; AAB06507.1; JOINED.
CC EMBL; A21577; CAA01560.1; -;
CC PIR; A35833; A35833.
CC PIR; A37816; A37816.
CC PIR; B37816; B37816.
CC PIR; A41386; A41386.
CC PIR; PLO135; PLO135.
CC PIR; PLO136; PLO136.
CC PIR; S04662; S04662.
CC PIR; S07433; S07433.
CC PIR; S09339; S09339.
CC PIR; S34056; S34056.
CC PIR; S43646; S43646.
CC SWISS-2DPAGE; P10909; HUMAN.
CC MIN; 185430; -;
CC InterPro; IPR000753; Clusterin.
CC Pfam; PF01093; Clusterin; 1.
CC SMART; SM00035; CLA; 1.
CC SMART; SM00030; CLB; 1.
CC PROSITE; PS00492; CLUSTERIN_1; 1.
CC PROSITE; PS00493; CLUSTERIN_2; 1.
CC Apoptosis; Complement pathway; Glycoprotein; Plasma; Sulfation;
Signal.
FT SIGNAL 1 22
FT CHAIN 23 449
FT CHAIN 23 227
FT CHAIN 228 449
FT DISULFID 102 313
FT DISULFID 113 305
FT DISULFID 116 302
FT DISULFID 121 295
FT DISULFID 129 285
FT CARBOHYD 86 86
FT CARBOHYD 103 103
FT CARBOHYD 145 145
FT CARBOHYD 291 291
FT CARBOHYD 354 354
FT CARBOHYD 374 374
FT CONFLICT 28 28
FT CONFLICT 47 47
FT CONFLICT 52 52
FT CONFLICT 305 305
CLUSTERIN.
BETA-CHAIN.
ALPHA-CHAIN.
INTERCHAIN.
INTERCHAIN.
INTERCHAIN.
INTERCHAIN.
N-LINKED (GLCNAC. . .).
N-LINKED (GLCNAC. . .).
N-LINKED (GLCNAC. . .).
N-LINKED (GLCNAC. . .).
N-LINKED (GLCNAC. . .).
D -> S (IN REF. 9 AND 10).
Q -> H (IN REF. 11).
G -> O (IN REF. 11).
C -> M (IN REF. 10).

```

Query Match      17.2%   Score 434;   DB 1;   Length 449;
Best Local Similarity 25.8%;   Pred. No. 2,2e-19;
Matches 124;   Conservative 96;   Mismatches 208;   Indels 52;   Gaps 14;

QY 12 MKPELLVIVCLLMKSHCAPTWK-----DKTAISENLSKFSEGEIDAEDEEVKKALT 65
DB 1 MMKTLFLVGLL-----TWESQVLGDQTVSDNELQEMSGKYVKNKEIQNAV 51
QY 66 GIKOMKTMERKEKHTNLMSTLKKCKEEOEALKLNEVQHEERLCRESLADSWG 125
DB 52 GVKQIKTLIEKTEERTLLSNLEAKKEDALNETRESETKLKLPGVCNETMMALWE 111
QY 126 ECRSCLENNCKMRYT-TCQPSWSSVKNKIRFERFKIYQFLPFPHEDNEKDLPISEKLEK 184
DB 112 ECKPCLQKTCMKFYARVCRSGSLVGRQLREELNQSFPYFWNNGDR-----IDSLEEN 165
QY 185 DAQLTOMEDV-----FSQLTVDVNSLFRNRFVFMQOQEDQTFQSHFIS-DTDLTPEYF 239
DB 166 DRQTHMLDVNQDHFSSRASSIIDLFQDRF-----FTREPQDTY--HYLPFSLPHRRPHF 218
QY 240 FPAFSKEPMTKADLE-QCWDIPNFQFCNFSVSIYESVSETITKMLKAIEDLPKQDKAP 298
DB 219 F--FPKSRIVSRSLMPFSPYEPFLPHANFPQPELEIHEA-----QQAMDIFHFSAPF 267
QY 299 DHGLLSKMLPGQDGLGCLGELDONLSRCFKFHEKCKQCAHLSEDC----PDVPAHLTEL 354
DB 268 QHPPTPEIREGDDRTVCREIRHNSTGCLRMKQCDCKREILSVDCSTNNPSQAKLRREL 327
QY 355 DEARLVNVSNOQYGOILQMTKHELTAVLVEKMRQFCGWSELANOAPETELIFNSIQ 414
DB 328 DESIQVAERTUTRIKELLKQYKMLNTSSLLBOLNBFQNVNVRSLANLTQGEDQYILRVT 387
QY 415 VVPRHIEGNISKQDETMTDLSILPSSNFTLKIPLESABESSNFIVGVVAKALQHF-KEH 473
DB 388 TVAS-HTSDSDVPSGVTEVVVKLFDSDPITVTPVEVSRKNPKFEMETVAEKALQETRKKH 446

RESULT 2
CLUS_COTJA STANDARD; PRT; 451 AA.
AC PI4018;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Clusterin precursor (51.5 kDa protein).
GN Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroretina;
RX MEDLINE=89239492; PubMed=2541393;
RA Michel D., Gillet G., Volovitch M., Pessac B., Calothy G., Brun G.;
RT "Expression of a novel gene encoding a 51.5 kD precursor protein is
RT induced by different retroviral oncogenes in quail neuroretinal
RT cells."
RL Oncogene Res. 4:127-136(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95262670; PubMed=7744033;
RA Michel D., Chatelain G., Herault Y., Brun G.;
RT "The expression of the avian clusterin gene can be driven by two
RT alternative promoters with distinct regulatory elements."
RL Eur. J. Biochem. 229:215-223(1995).
CC -!- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH.
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER

```

```

(BY SIMILARITY)
-!- INDUCTION: BY DIFFERENT RETROVIRAL ONCOGENES.
-!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DB EMBL: X15825; CAA33823.1; -
DB EMBL: X80760; CAA56733.1; -
DB PIR: S07714; S07714.
DB InterPro: IPR000753; Clusterin.
DB Pfam: PF01093; Clusterin; 1.
DB SMART: SM00035; Cln; 1.
DB SMART: SM00030; Cln; 1.
DB PROSITE: PS00492; CLUSTERIN_1; 1.
DB PROSITE: PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 451 CLUSTERIN.
FT CHAIN 19 230 BETA-CHAIN (SMALL SUBUNIT) (POTENTIAL).
FT CHAIN 231 451 ALPHA-CHAIN (LARGE SUBUNIT) (POTENTIAL).
FT DISULFID 98 314 INTERCHAIN (BY SIMILARITY).
FT DISULFID 109 306 INTERCHAIN (BY SIMILARITY).
FT DISULFID 112 303 INTERCHAIN (BY SIMILARITY).
FT DISULFID 117 296 INTERCHAIN (BY SIMILARITY).
FT DISULFID 125 286 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 355 355 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 437 437 A -> P (IN REF. 1).
FT CONFLICT 445 451 KONNTIE -> SRTTP (IN REF. 1).
SQ SEQUENCE 451 AA; 51800 MW; 8CF811225C5B22F3 CRC64;

Query Match      16.5%   Score 418;   DB 1;   Length 451;
Best Local Similarity 25.7%;   Pred. No. 2,1e-18;
Matches 118;   Conservative 94;   Mismatches 179;   Indels 68;   Gaps 15;

QY 45 LKSFSEGEIDAEVKKALTGKIKMKIMMERKEKHTNLMSTLKKCREBKQALKLNE 104
DB 27 LKLSAAGSKYIDAEVENAINGVKMTLMDKTSKEHQAMLHTLETKKKKEAVKALE 86
QY 105 VQHELEERLCRESLADSWGECRSCLENNCMRYT-TCQPSWSSVKNKIRFERFKIYOF 163
DB 87 KEKQALAEKQEVNCTMLSLWECKPCLKHTCMRVYSKMKCHSGSLVGRQLEELNRSSPF 146
QY 164 LPFFHEDNEKDLPISEKLEK-----DAQLTOMEDVFSQLTVDVNSLFRNRFVFMQOQ 218
DB 147 SIWNGERIDDDLLDREQQRERFEDLEERFGLMEDVEDIFQDSTQLYGPAPFFER---- 202
QY 219 EFDQTFQSHFISDTDLTEPY--FEPAFSKEPMTKADLEQCWDIPN-----FQLFC 267
DB 203 -----TPPFGGFEAF-VPPVORVHL-----VPRRRLSRELHPFFQHPM 240
QY 268 NFSVSIYESVSETITKMLKAIEDLPKQDKAPDH--GGLISKMLP-QQDRGLCGELQNLIS 324
DB 241 HGFHRLFPQLEFEMTQHLMDG-----HGAWHEPLGSGFATESRNFSTRMVCREIRNSA 294
QY 325 RCFKPFHEKCKQCAHLSEDC----PDVPAHLTELDEARLVNVSNOQYGOILQMTKHELE 380
DB 295 GCLRMDECEKREILAVDVCSDTPVQSOLREQFEDALRAERFTRRYDOLLSAFOAEML 354
QY 381 DTAYLVKMRQFCGWSELANOAP-----ETELIFNSIQVVRPIHEGNTSKODETMTD 434
DB 355 NTSSLLDQLNRQFGWVSRGLNLTQGNDFGLQVTTVFESK---TPNLEDP--SAPADQVVT- 408

```


(Apolipoprotein J) (Apo-J).
 CLU OR MSGP-2 OR APOJ.
 Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Testis;
 RC MEDLINE=93356785; PubMed=8352774;
 RA Lee K.-H., Ji Y.-M., Lim H.M., Lee S.-C., You K.-H.;
 RT "Molecular cloning and sequencing of sulfated glycoprotein-2 cDNA
 from testis of mouse: implications of two different mRNAs of SGP-2.";
 RL Biochem. Biophys. Res. Commun. 194:1175-1180(1993).
 [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=93359508; PubMed=8354695;
 RA French L.E., Chonn A., Ducrest D., Baumann B., Belin D., Wohlwend A.,
 RC Kisse J.Z., Sappino A.P., Tschopp J., Schiffrin J.A.;
 RT "Mouse clusterin: molecular cloning and mRNA localization of a gene
 associated with epithelial differentiation processes during
 embryogenesis.";
 RL J. Cell Biol. 122:1119-1130(1993).
 [3]
 RN SEQUENCE FROM N.A.
 RP Hodgdon B.A., Min B.H., Yan H., Farris J.A., Foster D.N.,
 RA Strauch A.R.;
 RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
 [4]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6 X CBA; TISSUE=Liver;
 RC MEDLINE=94223204; PubMed=8169523;
 RA Jordan-Starck T.C., Lund S.D., Witte D.P., Aronow B.J., Ley C.A.,
 RT Stuart W.D., Sverttger D.K., Clayton L.R., Sells S.F., Paigen B.;
 RT "Mouse apolipoprotein J: characterization of a gene implicated in
 atherosclerosis.";
 RL J. Lipid Res. 35:194-210(1994).
 CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A
 CC VARIETY OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS,
 CC MEMBRANES, AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED
 CC WITH PROGRAMMED CELL DEATH.
 CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
 CC -!- TISSUE SPECIFICITY: MOST ABUNDANT IN STOMACH, LIVER, BRAIN, AND
 CC TESTIS, WITH INTERMEDIATE LEVELS IN HEART, OVARY, AND KIDNEY.
 CC -!- PTM: EXTENSIVELY GLYCOSYLATED WITH SULFATED N-LINKED
 CC CARBOHYDRATES.
 CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D14077; BAA03162.1; -;
 CC EMBL; L08235; AAA37422.1; -;
 CC EMBL; L05670; AAA37284.1; -;
 CC EMBL; S70244; AAB30623.1; -;
 CC PIR; A40714; A40714.
 CC PIR; JN0699; JN0699.
 CC MGD; MGI:88423; Clu.
 CC InterPro: IPR000753; Clusterin.
 CC Pfam; PF01093; Clusterin; 1.
 CC SMART; SM00035; Clu; 1.
 CC SMART; SM00030; CLB; 1.
 CC PROSITE; PS00492; CLUSTERIN_1; 1.
 CC PROSITE; PS00493; CLUSTERIN_2; 1.
 CC Glycoprotein; Sulfation; Signal.
 FT SIGNAL 1 21 BY SIMILARITY.
 FT CHAIN 22 448 CLUSTERIN.
 FT BETA-CHAIN (SMALL SUBUNIT)
 FT CHAIN 22 226

Db 233 ----NFHDVFQPF-----YDHIHQQAQAMDAHLQ-----RTPVHFPTM-EFTENNDRTVC 277

QY 317 GELDONLSRCFKHEKQKQAHLSDC-----PDVPAHTEDEAIRLVNVNQYQOIL 372

Db 278 KEIRHSTGCLRWKQCKEKEQLEVDSCASNPTQTLRQOLNASOLAKEFSLRYDQLL 337

QY 373 QMTRKKHLEDYALVEMKRGFGWVSELANAPETE-----IIFNSQVVPRIHEGNISK 426

Db 338 QSYQOKMLNTSALLKQLEQFTWVSOLANTQSDQHYLVFTVNSHNSDPSIPSG----- 393

QY 427 ODEMTMTDLSILPSSNFTLPIPLEESNESSNFYGVVAKALQHF 471

Db 394 -----LTKVIVKLNFSPITVTVPQEVSSPNEFNAEKAQOQYR 433

RESULT 7

CLUS_PIG STANDARD; PRT; 446 AA.

AC Q29549;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Clusterin precursor (Complement cytolysis inhibitor) (CLI).

GN CLU.

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92184774; PubMed=1544909;

RA Diemer V., Hoyle M., Baglioni C., Millis A.J.;

RT "Expression of porcine complement cytolysis inhibitor mRNA in cultured aortic smooth muscle cells. Changes during differentiation in vitro."

RL J. Biol. Chem. 267:5257-5264(1992).

CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES, AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED CELL DEATH (BY SIMILARITY).

CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY SIMILARITY).

CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS IN BRAIN AND LIVER; LOWER LEVELS ARE DETECTED IN OTHER TISSUES, INCLUDING THE AORTA.

CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M84639; AAA31013.1; -

DR InterPro: IPR000753; Clusterin.

DR Pfam: PF01093; Clusterin; 1.

DR SMART: SM00035; CLA; 1.

DR SMART: SM00030; CLB; 1.

DR PROSITE: PS00492; CLUSTERIN_1; 1.

DR PROSITE: PS00493; CLUSTERIN_2; 1.

KW Glycoprotein; Signal.

FT SIGNAL 1 28

FT CHAIN 29 446

FT CHAIN 29 227

FT CHAIN 228 446

FT CHAIN 102 312

FT DISULFID 113 304

FT DISULFID 116 301

FT DISULFID 121 294

FT DISULFID 129 284

FT CARBOHYD 86 86

FT CARBOHYD 103 103

BY SIMILARITY.

CLUSTERIN.

BETA-CHAIN (A CHAIN).

ALPHA-CHAIN (B CHAIN).

INTERCHAIN (BY SIMILARITY).

INTERCHAIN (BY SIMILARITY).

INTERCHAIN (BY SIMILARITY).

INTERCHAIN (BY SIMILARITY).

INTERCHAIN (BY SIMILARITY).

N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 446 AA; 51774 MW; BLD5B434B668E3AA CRC64;

Query Match 15.8%; Score 400.5; DB 1; Length 446;

Best Local Similarity 24.4%; Pred. No. 2.3e-17;

Matches 118; Conservative 98; Mismatches 196; Indels 71; Gaps 16;

QY 16 LLVFIIVCLLWLKSDHCAPTWK-----DKTAISENLKSFSEVGEIDAEEVKALTGK 68

Db 4 LLLLVGILL-----TWENGPWVLGDKAISDKELQEMSTEGSKYVYVKEIKNALEK 54

QY 69 QMKIMMERKEKEHTNLMLSTLKKCRBEKQKALKNLNEVQEHLEERLCRESLADSGECR 128

Db 55 QIKTLIEQSNNEERKSLSSLEAAKKKEDALNDTRDTETKLGSGQLCNETMMALWEECK 114

QY 129 SCLENNCMRIYT--TCQPSHSSVKNKIERFKIYQFLFPFHEDNEKDLPISEKLIKDAQ 187

Db 115 PCLKQTCMKFYARVCRSGSLVGHQLEEFNLQSSPFYFWINGDR-----IDSLMENDRQ 168

QY 188 LTOMEDVFSLTVDVNSLENRSFNVFQMQEFDQFOSHFI-----DTDLTEPY----- 238

Db 169 QSHVMDI-----MEDSFNRASNI-----MDELQDRFFNREPFDFOFFSPFGSSHR 214

QY 239 ---FFPAFSKEPMTKADLEQCWDIPNFOLFNCFSIYESVSETITKMLKALDELPKOD 295

Db 215 GSLFFNPKSRFARNIMPFPLFTDL-NYHMFQPF---FDMIHQAOQAMDAHLRIPIYH- 268

QY 296 KAPDHGGLISKMLP--GQDRGLCGELDONLSRCFKHEKQKQAHLSDC-----PDVPA 349

Db 269 -FPEAG-----VPENSNDRAVCKEIRHNSGTCLRMKQCKEKEITLIPQEV--DPKFMETVAEALQ 321

QY 350 LHTDELDEAIRLVNVNQYQOILQMTKHELDYALVEMKRGFGWVSELANAPETEII 409

Db 322 LRQELYSLOMAEKFSKLYDQLLOYSQOQKMLNTSSLLKQLNEQFSWVSQLANTQNDYR 391

QY 410 FNSIQVVPRIHEGNISKQDETMWTDLSILPSSNFTLPIPLEESNESSNFYGVVAKALQ 469

Db 382 YLQVTTV-NSHGSDPSVPSGLTKVYVVKLFDSYPTILLIPEVS--DPKFMETVAEALQ 438

QY 470 EKE 472

Db 439 YRQ 441

RESULT 8

CLUS_RAT STANDARD; PRT; 447 AA.

AC P05371;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Clusterin precursor (Sulfated glycoprotein 2) (SGP-2) (Dimeric acid glycoprotein) (DAG) (Testosterone repressed prostate message-2) (TRPM-2).

DE 2).

GN CLU.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-45 AND 227-241.

RX MEDLINE=88000523; PubMed=3651384;

RA Collard M.W., Griswold M.D.;

RT "Biosynthesis and molecular cloning of sulfated glycoprotein 2 secreted by rat Sertoli cells."

RL Biochemistry 26:3297-3303(1987).

RN [2]

RP SEQUENCE FROM N.A.

RA Pineault J.M., Tenniswood M.;
 RN Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89149740; PubMed=2920020;
 RA Bettuzzi S., Hlipakka R.A., Gilna P., Liao S.;
 RT "Identification of an androgen-repressed mRNA in rat ventral prostate
 as coding for sulphated glycoprotein 2 by cDNA cloning and sequence
 analysis.";
 RL Biochem. J. 257:293-296(1989).
 RN [4]
 RP SEQUENCE OF 22-51 AND 227-256.
 RX MEDLINE=88326333; PubMed=3415696;
 RA Cheng C.Y., Chen C.C., Feng Z., Marshall A., Baird C.W.;
 RT "Rat clusterin isolated from primary Sertoli cell-enriched culture
 medium is sulfated glycoprotein-2 (SGP-2).";
 RL Biochem. Biophys. Res. Commun. 155:398-404(1988).
 RN [5]
 RP CHARACTERIZATION OF TRPM-2.
 RX MEDLINE=90134121; PubMed=2299741;
 RA Bandyk M.G., Sawczuk I.S., Olsson C.A., Katz A.E., Buttyan R.;
 RT "Characterization of the products of a gene expressed during
 androgen-programmed cell death and their potential use as a marker
 of urogenital injury.";
 RL J. Urol. 143:407-413(1990).
 CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
 OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
 AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
 CELL DEATH.
 CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED BY CELLS UNDERGOING PROGRAMMED
 DEATH AS A RESULT OF THE HORMONAL STIMULI OR A TRAUMATIC INSULT.
 CC -1- PTM: EXTENSIVELY GLYCOSYLATED WITH SULFATED N-LINKED
 CARBOHYDRATES.
 CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M16975; AAA41273.1; -;
 DR EMBL; M64723; AAA42298.1; -;
 DR EMBL; M64733; AAA42299.1; -;
 DR EMBL; X13231; CAA31618.1; -;
 DR PIR; A27205; A27205;
 DR PIR; A31575; A31575;
 DR PIR; B31575; B31575;
 DR PIR; A45890; A45890.
 DR PIR; S18491; S18491.
 DR InterPro: IPR000753; Clusterin.
 DR Pfam: PF01093; Clusterin; 1.
 DR SMART; SM00035; Clu; 1.
 DR SMART; SM00030; Clb; 1.
 DR PROSITE; PS00492; CLUSTERIN_1; 1.
 DR PROSITE; PS00493; CLUSTERIN_2; 1.
 KW Sulfation; Glycoprotein; Spermatogenesis; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 447 CLUSTERIN.
 FT CHAIN 22 226 BETA-CHAIN (SMALL SUBUNIT).
 FT CHAIN 227 447 ALPHA-CHAIN (LARGE SUBUNIT).
 FT DISULFID 101 312 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 112 304 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 115 301 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 120 294 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 128 284 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 102 102 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (PROBABLE).

FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CONFLICT 187 187 D -> H (IN REF. 1).
 SQ SEQUENCE 447 AA; 51375 MW; 9E2FA33E5E0C146E CRC64;
 Query Match 15.0%; Score 380; DB 1; Length 447;
 Best Local Similarity 23.5%; Pred. No. 3.9e-16;
 Matches 113; Conservative 103; Mismatches 197; Indels 68; Gaps 15;
 QY 16 LLFIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEID-----ADEYVKKALTGIKQ 69
 DB 4 LLLCVALL-----TWNGWVGLGEQESDNELQELSTQGSRYVYNKELQNAVQGVKH 54
 QY 70 KIMMERKEKHTNLMSTLKKREEKQALKLNEVQHLKEERLCRESLADSGEGRS 129
 DB 55 IKTLIETNAERKILLSLEAKKKKGGALDDTRDSEMKLKAFFVCNCTMMALWEECKP 114
 QY 130 CLENNCMRIYT-TCOPSWSSVYKNKIERFRKIYQFLPFFHEDNEKDLPISEKLLIEKDAQL 188
 DB 115 CLKHTCMKFYARVCRSGSLVGRLEEFNLQSSPFYFMMNGDR-----IDSLLESDRQQ 168
 QY 189 TQ-----MEDVFSQLTVDVNSLNFNSFNVRQMQQEFQDTFQSHFISDTDLTEPYFFP--- 241
 DB 169 SQVLDAMQDSFRASGIIDTLFQDRF--FTHPEQDI-----HHF-----SPWGFPHKR 214
 QY 242 ---AFSKEPTMKADLEQCWDIP-NFFQLFCNFSVSVSVSETITTKMLKAIEDLPKQDKA 297
 DB 215 PHFLYPKSRLVSLMPLSHYGLPSLFHNMFPFFDHIHOAQOAMDVLHSPALQFPDVF 274
 QY 298 PDHGLISKMLPGQ-DRLCGELQNLSCRFKFEKCKQCOAHLSDEC-----PDVPALHT 352
 DB 275 KE-----GEDDPTVCKEIRHNSGCLKMKGCKEKEILSVDCSTNNPQAANLRQ 324
 QY 353 ELDEAIRLVNSQYGOILQWTRKHLDTAYLVKMRGQFCGWSELAN--QAPETEIRF 410
 DB 375 ELNDSLOVAERLTQYNNELLHSLQSKMLNTSSLLEQLNDQFTWVSQLANLTQGDQYLRV 384
 QY 411 NSIQVVPRIHGNTSKQDETMMTDLISLPSSNFTLKIPLESASNSNFIGYVAKALQHF 470
 DB 385 STVTT---HSSDSVPSRVTEVVVVKLFDSDPITVVLPEEYVSKNPKFMDTVAEKALQEY 440
 QY 471 K 471
 DB 441 R 441

RESULT 9
 CLUS_MESAU
 ID CLUS_MESAU STANDARD; PRT; 191 AA.
 AC P14683;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DE 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Clusterin (Sulfated glycoprotein 2) (SGP-2) (Fragment).
 GN CLU.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE OF 24-123 FROM N.A.
 RA MEDLINE=89386721; PubMed=2780570;
 RX Duquid J.R., Bohmont C., Liu N., Tourtellotte W.W.;
 RT "Changes in brain gene expression shared by scrapie and Alzheimer
 disease.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:7260-7264(1989).
 RN [2]
 RP SEQUENCE OF 1-23 AND 124-191 FROM N.A.
 RX Duquid J.R., Bohmont C., Liu N., Tourtellotte W.W.;
 RL Submitted (AUG-1989) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY

CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH.
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC -1- PTM: EXTENSIVELY GLYCOSYLATED WITH SULFATED N-LINKED
CC CARBOHYDRATES.
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M26640; AAA37102.1; -
CC InterPro: IPR000753; Clusterin.
CC Pfam: PF01093; Clusterin; 1.
CC SMART: SM00035; Cla; 1.
CC PROSITE: PS00492; CLUSTERIN_1; PARTIAL.
CC PROSITE: PS00493; CLUSTERIN_2; 1.
CC Sulfation; Glycoprotein; Spermatogenesis.
CC NON_TER 1
CC CHAIN <1 >191 CLUSTERIN.
CC CHAIN <1 >15 BETA-CHAIN (SMALL SUBUNIT).
CC CHAIN 16 >191 ALPHA-CHAIN (LARGE SUBUNIT).
CC FT CARBOHYD 79 79 N-LINKED (GLCNAC. .) (PROBABLE).
CC FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (PROBABLE).
CC FT NON_TER 191 191
CC SEQUENCE 191 AA; 22099 MW; 20A9E7E6963C619 CRC64;

Query Match 6.6%; Score 168; DB 1; Length 191;
Best Local Similarity 35.1%; Pred. No. 0.00083;
Matches 33; Conservative 22; Mismatches 35; Indels 4; Gaps 1;
QY 312 DRGLGELDONLSRCFHEKCKQKQAHLSDEC-----PDVPAHLTEDEAIRLVNVSQ 367
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 69 DRAVCKEIRHNSGCLMKMGCKECCILSVDCSANNPAQHLRQLNDSLQVAERLTQR 128
QY 368 YGQILQWTRKHLDEYALVEKMRGQGWSELAN 401
DB 129 YNELLHSLOTKMLNTSLLEQLNEQFNWVSQLAN 162

RESULT 10
MYS2_SCHPO STANDARD; PRT; 1526 AA.
AC Q9US16; P78969;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin type II heavy chain 1.
GN MYO2 OR SPC645.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=972;
RX MEDLINE=98075862; PubMed=9415380;
RA May K.M., Watts F.Z., Jones N., Hyams J.S.;
RT "Type II myosin involved in cytokinesis in the fission yeast,
RT Schizosaccharomycetes pombe";
RL Cell Motil. Cytoskeleton 38:385-396(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Rieger M.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: REQUIRED FOR CELL DIVISION. IT IS A COMPONENT OF THE
CC CDC12 'SPOT', A STRUCTURE THOUGHT TO MARK THE SITE OF SEPTATION.
CC MAY WORK IN CONJUNCTION WITH MYO3.
CC -1- SUBUNIT: BINDS TO CDC4 AND RLC1.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U75357; AAC49908.1; -
CC EMBL: AL049498; CAB39901.1; -
CC HSSP: P08799; LMND.
CC InterPro: IPR000048; IQ.
CC InterPro: IPR004009; Myosin_N.
CC InterPro: IPR001609; myosin_head.
CC Pfam: PF00612; IQ; 1.
CC Pfam: PF00063; myosin_head; 1.
CC Pfam: PF02736; Myosin_N; 1.
CC PRINTS: P00193; MYOSINHEAVY.
CC ProDom: PD000355; myosin_head; 1.
CC SMART: SM00015; IQ; 1.
CC SMART: SM00242; MYSC; 1.
CC PROSITE: PS00096; IQ; 1.
CC Myosin; Actin-binding; ATP-binding; Coiled coil; Calmodulin-binding;
KW Alkylation.
KW DOMAIN 1 757 MYOSIN HEAD-LIKE.
FT DOMAIN 758 787 IQ.
FT DOMAIN 875 1244 COILED COIL (POTENTIAL).
FT NP_BIND 177 177 ATP (POTENTIAL).
FT DOMAIN 634 656 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 734 748 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 674 674 ALKYLATION (BY SIMILARITY).
FT CONFLICT 1337 1337 S -> R (IN REF. 1).
SQ SEQUENCE 1526 AA; 176430 MW; D71D51D6578192BA CRC64;

Query Match 6.0%; Score 151; DB 1; Length 1526;
Best Local Similarity 20.6%; Pred. No. 0.0095;
Matches 98; Conservative 73; Mismatches 198; Indels 106; Gaps 17;
QY 47 SPSEVGEIDADEVKKALTGKIMMERKEKHTNLMSTLKKCKREKQKALKLLNEVQ 106
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 895 SFSETKQ--QNEINLQRESASLKQINNELESELLEKTSKVETL-----LSEQNELK 942
QY 107 EHLEEEERLCRESIADSWGECRSCLENNCMRIYTCOPSSSVKNKIERFRKIYQLFP 166
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 943 EKUSLEEK-----DLDTKGELESLENN-----ATVLSKAEF----- 976
QY 167 FHEDNEKDLPISEKLEKDAQLTQMEDVFSQLTVDVNS--LFNSFNVFMQOQED--- 221
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 977 ----NQCKSLQETIVTKDAELDKLKYISDYKTEIQEMRLTNQKME-KSIQOEGSLSE 1031
QY 222 -----QTFQSHFTISDTDLTEP-----YFFAFSKEPMT 249
DB 1032 SLKRVKKLERENSTLSDVSLKQKEELSVLKGQVELTINNLEEKVNYLEADVQLPKL 1091
QY 250 KADLEQCWDIPNFFOLFNCFSVSIYSVETITKMLKAIEDLP-KQDKADPHGLGLSKML 308
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 1092 KREESLNDKQQLYQLOATKNKELEAKVKECLNNIKSLTKLELENKEEKCONLSASLYI 1151
QY 309 PGDRLGCGELDONL-----SRCKFKHEKCKQKQAHLSDECPCDVPALHTEDEAIRLVNS 364
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 1152 ELQ-----EIHENLLLVSDLENYKKVEGLQLDL-EGLKVDVTFQELSKKHRDLTFN 1204
QY 365 NQ-----QYGOILQWTRKHLDEYALVEKMRGQGWSELANQAPETEIFNSIQVVP 417
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 1205 HESLLRQSASYKEKLSLASSENKDLNKNVSSLTQKNELSPKASKVPELE-----RKITN 1259

```
QY 418 RIHE-GNISKQDETMMTDLSPSSNFTLKIPLESASSNFYGVVAKALQHFK 471
ID C1K1_YEAST STANDARD: PRT; 594 AA.
Db 1260 LMHEYSQLGKTFEDEKRALIASRONEELR-SLKSELESKRKLEVEYQKVLEEVK 1313

RESULT 11
C1K1_YEAST
AC Q01649;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Spindle POLE body associated protein.
GN C1K1 OR YMR198W OR YMR946.11.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92354905; PubMed=1644287;
RA Page B.D., Snyder M.;
RT "C1K1: a developmentally regulated spindle pole body-associated
RT protein important for microtubule functions in Saccharomycetes
RT cerevisiae.";
RL Genes Dev. 6:1414-1429(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: A DEVELOPMENTALLY REGULATED PROTEIN IMPORTANT FOR
CC MICROTUBULE FUNCTIONS. TIGHTLY ASSOCIATED WITH KAR3, MAY SERVE TO
CC REGULATE THE CELLULAR COMPARTMENT IN WHICH KAR3 FUNCTIONS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M96439; AAA34494.1; -.
DR EMBL; Z47815; CAA87820.1; -.
DR PIR; A44073; A44073.
DR SGD; S0004811; C1K1.
KW Microtubules.
SQ SEQUENCE 594 AA; 69069 MW; A2A09DEB8C2EC838 GRC64;

Query Match 5.8%; Score 145.5; DB 1; Length 594;
Best Local Similarity 20.9%; Pred. No. 0.069;
Matches 84; Conservative 71; Mismatches 138; Indels 109; Gaps 17;

QY 26 LKDSHCAPTKWKTAKISENLKSFVSG-----BIDAEVYKKA----- 63
Db 147 LKDNQCDLQRKVELSSKNIVSMHAKVQVEFENDELEELSNAKREWTYKLMVENLKPDE 206
QY 64 -LTG-IKQWKIMWE--RK-----EKEHNTLMSTLTKCKREEKQAL-KLLNEVQEHLE-EE 112
Db 207 RLTDENRQLKTEFEVNRKRLFLOQENENECKNYKELDKKEIFPKVKVNDARIELDGEQ 266
QY 113 ERLCR--ESLADSWGECRSLNNCMRIYTTQPSWSSYKNNKIRFFRKIYQFLFPFFHED 170
Db 267 ERLSKVLKDLQDTHGELKENIK-----TC-----RDEFNDFEKRIGEAEVNFHSM 311
QY 171 NEKDLPISEKLEKDAQLTQMEDVFSQLTVDVNSLNFNSFNVRMQQEFDTFOSHPIIS 230
Db 312 ELAVVPLKKKLASTQALTVQVEERKQVGEAGANNKKIYVNELEKVOQEL----- 361
QY 231 DTDLTPEYFFPAFSEKPEMTKADLE-----QCWDIPNFFOLFPCNFVS-----IYESVSET 280
```

```
Db 362 -----YTRQNLATSTIEEIKGYTRCFAYANERQMPDEFHINXVDRICENSGE- 408
QY 281 ITKMLKATEDLPKQDKAPDHGGLISKMLPQGDRLGCLGELDONLSRCFKFHEKCKQKQAH 340
Db 409 --KRQVQFDRVVLEIHKDKHRLYNECIPFLE-----KYISKLNCSIIV 451
QY 341 SEDCPDVPALHTELDEAIRLVNSNQYQQLQWTRK--HLE 380
Db 452 VSQOPTAPMKKTKLLKQLI-----EQYGENYKMTLNLHLHD 486

RESULT 12
MYHD_HUMAN STANDARD: PRT; 1938 AA.
ID Q9UKX3; O95252;
AC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, skeletal muscle, extraocular (MyHC-ec).
GN MYH13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Extraocular muscle;
RX MEDLINE=99318869; PubMed=10388558;
RA Weiss A., Schiaffino S., Leinwand L.A.;
RT "Comparative sequence analysis of the complete human sarcomeric myosin
RT heavy chain family: implications for functional diversity.";
RL J. Mol. Biol. 290:61-75(1999).
RN [2]
RP SEQUENCE OF 1917-1938 FROM N.A.
RC TISSUE=Extraocular muscle;
RX MEDLINE=99026150; PubMed=9806854;
RA Winters L.M., Briggs M.M., Schachat F.;
RT "The human extraocular muscle myosin heavy chain gene (MYH13) maps to
RT the cluster of fast and developmental myosin genes on chromosome 17.";
RL Genomics 54:188-189(1998).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE SI DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF111782; AAD29948.1; -.
DR EMBL; AF075248; AAC83241.1; -.
DR HSSP; P08799; IMND.
DR MIM; 603487; -.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
```

DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00612; IQ; 2.
 DR Pfam: PF00063; myosin_head; 1.
 DR Pfam: PF02736; Myosin_N; 1.
 DR Pfam: PF01576; Myosin_tail; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR PRODOM: PD000355; myosin_head; 1.
 DR SMART: SM00015; IQ; 1.
 DR SMART: SM00242; MYSC; 1.
 DR PROSITE: PS00096; IQ; 1.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
 KW Multigene family.
 FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
 FT DOMAIN 785 814 IQ.
 FT DOMAIN 843 1938 COILED COIL (POTENTIAL).
 FT NP_BIND 179 186 ATP (POTENTIAL).
 FT DOMAIN 659 681 ACTIN-BINDING (BY SIMILARITY).
 FT DOMAIN 761 775 ACTIN-BINDING (BY SIMILARITY).
 FT MOD_RES 130 130 METHYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 699 699 ALKYLATION (SH-2) (POTENTIAL).
 FT MOD_RES 709 709 ALKYLATION (SH-2) (POTENTIAL).
 SQ SEQUENCE 1938 AA; 223678 MW; 1F6D006416381CD5 CRC64;

Query Match 5.88; Score 145.5; DB 1; Length 1938;
 Best Local Similarity 20.7%; Pred. No. 0.27;
 Matches 100; Conservative 93; Mismatches 206; Indels 83; Gaps 17;

QY 41 ISEMLKSFVEIGDADDEY-KKALTIKIMMERKEHEHTNLMSTLKKCR-----EKK 95
 DB 1150 ISEMLEAS--GATSAQIENKKEAFQKWRDLEATLOHEATATLRKQADSVAE 1207
 QY 96 OEALKLLNEVOEHLLEERLCRESLAD--SWGECRSCLENNCMRIYTTQPSWSSVKNKI 153
 DB 1208 GEQIDNLRVQKLEKESEKMEIDDMASNEALSCKSNIECTCTVEDQSEIKAKD 1267
 QY 154 ERFRKIYQFLFFPHEDNEKDLPISEKLEKDAQLTQMEDVFSQLVNVDNSLFRSNFV 213
 DB 1268 EQQTQI-----HLMNOKARLQQTQNGELSHRVEEKESLISQLTQKQALTOQLEELK 1320
 QY 214 ROMQOQFD-QTFQSHFIS-----TDLTEPYFFPAFKEPMTKADLBQCV-----DIPNFF 263
 DB 1321 ROMEETKAKNAHAHALQSSRHDCDLLR---EQVEEQEAKAEQALQALSKANSEVAQWK 1376
 QY 264 QLCNFSVSVIYESVETITKMLKAIEDL-PKQKADPHGGLISK-----MLPGQDRGLGCEL 319
 DB 1377 TKYETDAIQRTTELEEAQKLAQRLQEAEBEKTETANSKCSLEKTKORLOGEVEDLMRDL 1436
 QY 320 QDMLSRC-----PFHEKCKQCAHLSEDQDVPALHTDELDAIRLVNVS 365
 DB 1437 ERSHTACATLDKQRFNFKVLAQKQRLDESQAELEAAQKESRSSTEL---FKMRNAYE 1493
 QY 366 QYQGI--IQMTKRHLEDYALVLEKMRGFGWVSELANQAPET-----ELI 409
 DB 1494 EYVDQLETLRRENKNLOEE-----ISDLTEQIAETGNMLOEAKTKKLVQEO 1540
 QY 410 ENSIQVVRPHENISQDQETWM---TDLISLPSSNFTLIKPLESAESSNFIQGVVAKA 466
 DB 1541 KSDIQVALEVEGSLHEESKILRVOLLSQVKSLEDRKRVIEKDEIEQLKRNQRAAEA 1600
 QY 467 LQ 468
 DB 1601 LQ 1602

RESULT 13
 ID RA50_METJA
 AC Q58718;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE DNA double-strand break repair rad50 ATPase.
 GN RAD50 OR MJ1322.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A. 2561 / ATCC 43067;
 RC STRAIN=JAL-1 / DSM 2561 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bull C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii";
 RL Science 273:1058-1073(1996).
 CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
 CC rad50/mre11 complex possesses single-strand endonuclease activity
 CC and ATP-dependent double-strand-specific exonuclease activity.
 CC Rad50 provides an ATP-dependent control of mre11 by unwinding
 CC and/or repositioning DNA ends into the mre11 active site (By
 CC similarity).
 CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U67572; AAB99331.1; -
 CC TIGR: MJ1322; -
 DR InterPro: IPR003439; ABC_transportr.
 DR InterPro: IPR001687; ATP_GTP_A.
 DR InterPro: IPR003405; SMC_C.
 DR InterPro: IPR003395; SMC_N.
 DR Pfam: PF02483; SMC_C; 1.
 DR Pfam: PF02463; SMC_N; 1.
 KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
 FT NP_BIND 32 39 ATP (BY SIMILARITY).
 FT DOMAIN 158 849 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1005 AA; 119387 MW; 9BBB48173E788F3 CRC64;

Query Match 5.68; Score 141.5; DB 1; Length 1005;
 Best Local Similarity 20.0%; Pred. No. 0.22;
 Matches 81; Conservative 70; Mismatches 131; Indels 123; Gaps 16;

QY 36 KDKTAISENL-----KSFSEGEIDADEVKALGIKQIMMERKEHEHTNLMST 87
 DB 419 QEKSKIERNINDLEPRINKLLEETKNIDI-ESIENSLKEIEKKVLENKELKNNK 477
 QY 88 LKCKREEQKALKLLNEVOEHLLEERLCRESLADSWGECRSL-----ENNCMRIYTTQ 143
 DB 478 LGEINSE-----IKRLKILDELKEVE-----GKCPCKTTPIDENKKMELINQK 522
 QY 144 PSWSVKNKIERFRKIYQFLFFPHEDNEKDLPISEKLEKDAQLTQMEDVFSQLVNVD 203
 DB 523 TQLNNKYTELEINKKIREI-----EKDIEKLKEIDKEENLKTLLYLEKQSQIE 574
 QY 204 SLFNRSNFVRMQOQEDQTFQSHFISDITLTPYFFPAFKEPMTKADLBQCVQDIPNFF 263
 DB 575 EELKLLKN-YKEQLDEINKISNVYNGKPVDE-----ILE---DIKSQL 615
 QY 264 QLCNF-----SVSIYESVSET-ITKMLKAIEDLPKQKADPHGGLISKMLPGQDRGLC 316

```

Db 616 NKFNFEYNOYLSAVSYLNSVDEGIRNRKEIENI-----VSGWKK--- 656
Qy 317 GELDONLSRCKFKHEKQKQKQAHLSDCPDVPAHTELDE-----AIRLVNSNQ----- 366
Db 657 -----EKREELNKLREIREINRLKDKLNELKKEKELIEIENRRSLKFD 702
Qy 367 QYGOILOMTRKHELDYALVYKMRGQFGWVSELANAPETETIFN 411
Db 703 KYKEYLGLTEK-----LEELKNKIDGLEIYN 729

RESULT 14
UTRO_HUMAN
ID AC P46939; STANDARD; PRT; 3433 AA.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE utrophin (Dystrophin-related protein 1) (DRP1) (DRP).
CN UTRN OR DMDL
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93096045; PubMed=1461283;
RA Tinsley J.M., Blake D.J., Roche A., Fairbrother U., Riss J.,
RA Byth B.C., Knight A.E., Kendrick-Jones J., Suthers G.K., Love D.R.,
RA Edwards Y.H., Davies K.E.;
RT "Primary structure of dystrophin-related protein.";
RL Nature 360:591-593(1992).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 144-261;
RX MEDLINE=99141377; PubMed=9887274;
RA Keep N.H., Norwood F.L.M., Moores C.A., Winder S.J.,
RA Kendrick-Jones J.;
RT "The 2.0-A structure of the second calponin homology domain from the
RT actin-binding region of the dystrophin homologue utrophin.";
RL J. Mol. Biol. 285:1257-1264(1999).
CC -|- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE
CC PLASMA MEMBRANE (BY SIMILARITY TO DYSTROPHIN).
CC -|- SUBCELLULAR LOCATION: NEUROMUSCULAR JUNCTION.
CC -|- TISSUE SPECIFICITY: MUSCLE.
CC -|- SIMILARITY: STRONG, TO DYSTROPHIN.
CC -|- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
CC ABP-120, ABP-180, OR BETA-FODRIN).
CC -|- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -|- SIMILARITY: CONTAINS 1 WW DOMAIN.
CC -|- SIMILARITY: CONTAINS 20 SPECTRIN REPEATS.
CC -|- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X69086; CAA48829.1;
DR PIR; S28381; S28381.
DR PDB; 1BHD; 16-FEB-99.
DR NIM; 128240;
DR InterPro; IPR001589; Actinin_act_bind.
DR InterPro; IPR001715; Calponin_hom.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR003122; TarH.
DR InterPro; IPR001202; WW.
DR InterPro; IPR002349; WW_domain.
DR InterPro; IPR000433; ZnF_Z2.

```

```

DR Pfam; PF00307; CH; 2.
DR Pfam; PF00435; spectrin; 19.
DR Pfam; PF00397; WW; 1.
DR Pfam; PF00569; Z2; 1.
DR PRINTS; PR00403; WWDOMAIN.
DR SMART; SM00033; CH; 2.
DR SMART; SM00150; SPEC; 18.
DR SMART; SM00319; TarH; 1.
DR SMART; SM00456; WW; 1.
DR SMART; SM00291; ZnF_Z2; 1.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00020; ACTININ_2; 1.
DR PROSITE; PS00021; CH; 2.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PROSITE; PS00020; WW_DOMAIN_2; 1.
DR PROSITE; PS01357; ZF_Z2_1; 2.
DR PROSITE; PS0135; ZF_Z2_2; 1.
KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
KW Repeat; 3D-structure; Zinc-finger.
FT DOMAIN 1 246 ACTIN-BINDING.
FT DOMAIN 31 135 CH 1.
FT DOMAIN 150 252 CH 2.
FT REPEAT 253 308 SPECTRIN 1.
FT REPEAT 309 417 SPECTRIN 2.
FT REPEAT 418 526 SPECTRIN 3.
FT REPEAT 541 637 SPECTRIN 4.
FT REPEAT 687 798 SPECTRIN 5.
FT REPEAT 803 902 SPECTRIN 6.
FT REPEAT 1016 1083 SPECTRIN 7.
FT REPEAT 1125 1230 SPECTRIN 8.
FT REPEAT 1248 1334 SPECTRIN 9.
FT REPEAT 1432 1541 SPECTRIN 10.
FT REPEAT 1544 1649 SPECTRIN 11.
FT REPEAT 1652 1753 SPECTRIN 12.
FT REPEAT 1910 1968 SPECTRIN 13.
FT REPEAT 1976 2081 SPECTRIN 14.
FT REPEAT 2258 2333 SPECTRIN 15.
FT REPEAT 2399 2440 SPECTRIN 16.
FT REPEAT 2443 2556 SPECTRIN 17.
FT REPEAT 2559 2636 SPECTRIN 18.
FT REPEAT 2658 2688 SPECTRIN 19.
FT REPEAT 2691 2797 SPECTRIN 20.
FT DOMAIN 2812 2845 WW.
FT ZN_FING 3064 3111 Z2-TYPE.
SQ SEQUENCE 3433 AA; 394488 MW; EAE8DB409F858E5B CRC64;

Query Match 5.6%; Score 141.5; DB 1; Length 3433;
Best Local Similarity 22.2%; Pred. No. 0.9;
Matches 94; Conservative 71; Mismatches 176; Indels 83; Gaps 18;

Qy 27 KDSHCAPTWKDKTAISENLKSFSEVEIDADEEVKKALTGKQKIMMERKEKEHTNLS 86
Db 1551 KEAASLSWLSATETELVQKSTSEGLLDLDEISWAKNVKLDLE-----KKRADLNTIT 1605
Qy 87 TLKCKREKQKALKLLNEVQHLREELRESLADSGCRSLNENCMRIYITTCPSW 146
Db 1606 -----ESSAALQNLEGSPI-LEERLC--VLNAGHSRVRTWTEDWCNTL----- 1647
Qy 147 SSVKNKIERFRKI-----YQFLFPFHEDNEKDLPISEKLEIK-----DAQLTOMED 193
Db 1648 MNHQNLQLEIFDGNVAHISTWLYQAEALLDELEIKKPTSQEETVAKRLVSELDAHL-QVEN 1706
Qy 194 VFSQUTVDVNSLNFNRSNFVFMQOEFTQ--SHFISDLDLTPEYFPFAFSKEPTWKA 251
Db 1707 VRDQALILMNARGSSRELVEPKLAELNRNFKVSKHKSAKL-----LIAQEP----- 1755
Qy 252 DLEQCQDIPNPFOLFPCNFSVSIVESVETITKMLKAI-----EDLPKQDKAPDHGGLSKM 307
Db 1756 -LYQCLVTVTETETGVPF--SDLEKLENDIENMLKFEVKEHLESDDEKMDSESAQIEEV 1812
Qy 308 LPQDGRGLCGELDQNLRSRCFKFHEKQKQAHLSDCPDVPAHTELDEAIRLVNSNQ 367

```

Db 1813 LQGEEMLHOPMEDN-----KKEKIRLQLL-----LLHTRYNK-IKAIPQORK 1855

QY 368 YGQILQMTKRHLEDYALVEMKRGQFGWSELANQAPETELIENSIVQVPRHEGNLSKO 427

Db 1856 MGQLASIRSLPTDYLVE-----INKILLCDDVLSLNPVLPETAIYE-DFSQ 1906

QY 428 DETM 431

Db 1907 EDSL 1910

RESULT 15

CENE_HUMAN STANDARD; PRT; 2663 AA.

AC Q02224;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Centromeric protein E (CENP-E protein).

GN CENPE.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93024922; PubMed=1406971;

RA Yen T.J., Li G., Schaar B.T., Szilak L., Cleveland D.W.;

RT "CENP-E is a putative kinetochore motor that accumulates just before

RT mitosis";

RL Nature 359:536-539(1992).

RN [2]

RP CHARACTERIZATION.

RX MEDLINE=95196755; PubMed=7889940;

RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;

RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed

RT microtubule motor";

RL EMBO J. 14:918-926(1995).

RN [3]

RP CHARACTERIZATION.

RX MEDLINE=98437347; PubMed=9763420;

RA Chan G.K.T., Schaar B.T., Yen T.J.;

RT "Characterization of the kinetochore binding domain of CENP-E reveals

RT interactions with the kinetochore proteins CENP-F and hBUBR1.";

RL J. Cell Biol. 143:49-63(1998).

CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE

CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE

CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT

CC AND/OR SPINDLE ELONGATION.

CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.

CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING

CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS

CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.

CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Z15005; CAA78727.1; -

DR PIR; S28261; S28261.

DR HSSP; P17119; 3KAR.

DR MIM; 117143; -

DR InterPro; IPR001752; kinesin.

DR Pfam; PF00225; kinesin; 1.

DR PRINTS; PRO0380; KINESINHEAVY.

DR SMART; SM00129; KISC; 1.

DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.

DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.

KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;

KW Cell cycle; Centromere.

FT DOMAIN 1 335 KINESIN-MOTOR.

FT DOMAIN 336 2471 COILED COIL (POTENTIAL).

FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).

FT NP_BIND 86 93 ATP (BY SIMILARITY).

SQ SEQUENCE 2663 AA; 312087 MW; CEFC13880C8C8B8 CRC64;

Query Match 5.5%; Score 139.5; DB 1; Length 2663;

Best Local Similarity 18.1%; Pred. No. 0.88;

Matches 100; Conservative 87; Mismatches 175; Indels 191; Gaps 23;

QY 41 ISENKASFSEVGEIDADEEVKKALTGK-----OMKIMMERKEHEHNMSTLK----- 89

Db 1461 LKENKEI-VAKHLETEELKVAHCLKEQEBETINLRVNLSEKETEISTIQKLEAIND 1519

QY 90 ----KCRE--EKQEA--KILNEVQEHLE-----EERLCRESLADSWGECRSCLENNCM 136

Db 1520 KLONKIQEIYEKEEQLNIQISEQVENVNELKQFKEHRAKADSALOS-----IESKML 1572

QY 137 RYTTCPQSSSVK-----NRIERFFPKIYQFL- 164

Db 1573 ELTNRLQESQEEIQIMKEEMKRVQEAQIERDQLKENTKEIVAKMKESQEKYQFLK 1632

QY 165 -PPFHEDNEKDLPISEKLEKDAQLTQMEDVFSQLTVDVNSLNFNSFNVFRQMQEFQDT 223

Db 1633 MTAVNETQKMCIEHLEKQFQKLENLEIETE-NIRLTQILHENLEEMRSVTKERD-- 1689

QY 224 FQSHFISDLDLTPYFPFAPFSKEPMTKADLEQCWDIPNFFQFQFCNFSVSIYESVSTWK 283

Db 1690 -----DLR-----SVEETLKVERDQ-----LKENLRETIYR 1715

QY 284 MLKAIEDLPKQDK-----APDHGGLISKMLPGQDQRGGLGGLDQNLNRC----- 326

Db 1716 -----DLEKQOELKIVHMHLEHETIDKL-----RGIVSEKTNISNMOKDLEHSNDA 1764

QY 327 -----FKFHEKQCKQAHLSDCPDVPAL-----HTELDEAIRL 360

Db 1765 LKAQDLKIQEELRIAHMHLEKEQOETDKLRGIVSEKTDKLSNMOKDLENSNAKLEKIOE 1824

QY 361 VVNSQOYQOILQMTKRHLEDYALVEMKRGQFGWSELANQAPETELIENSIVQV----- 415

Db 1825 LKANHQ-----LITLKKDVNETQKKVSEM-----EQLKKQIKDQSLTSLKLEIENLNL 1873

QY 416 VPRIHEG-----NISKQDETMTDLSILPSSNFTLKIPLEESAESNFTGYV 463

Db 1874 AOELHENLEEMKSVKERNLRRVEETLKERD-----OLKESLQETKARDLEIQOEL 1926

QY 464 AKALQHFKHEHFT 476

Db 1927 KTARMLSKHEKET 1939

Search completed: July 2, 2002, 11:58:16

Job time: 369 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:57:10 ; Search time 53.04 Seconds
(without alignments)
1555.780 Million cell updates/sec

Title: US-09-722-544A-4MOD
Perfect score: 2529.
Sequence: 1 MRTWDSNSGNMKPPLLVFI.....FIGYVAKALQHFKEHFTW 477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2460	97.3	466	4 Q15846	Q15846 homo sapien
2	1832.5	72.5	465	6 Q93KN1	Q95knl canis famli
3	1226.5	48.5	338	6 Q9NIT8	Q9nit8 canis famli
4	423	16.7	449	6 Q29482	Q29482 equus cabal
5	402.5	15.9	448	13 Q9YGP0	Q9ygp0 gallus gall
6	321	12.7	372	11 Q9JK98	Q9jk98 mus musculu
7	266.5	10.5	295	11 Q35510	Q35510 rattus norv
8	204	8.1	218	11 Q76329	Q76329 dictyosteli
9	164	6.5	1738	5 Q9W6V0	Q9w6v0 gallus gall
10	161	6.4	3616	13 Q9W6V0	Q9w6v0 gallus gall
11	150	5.9	1156	16 Q66878	Q66878 aquifex aeo
12	149.5	5.9	840	6 Q95JRO	Q95jro macaca fasc
13	149	5.9	1388	6 Q28021	Q28021 bos taurus
14	148.5	5.9	1091	5 Q18082	Q18082 caenorhabdi
15	147	5.8	3259	4 Q14789	Q14789 homo sapien
16	145	5.7	1388	4 Q75116	Q75116 homo sapien

17	145	5.7	1388	4 Q9UQNS	Q9uqn5 homo sapien
18	144	5.7	1379	11 Q62868	Q62868 rattus norv
19	143	5.7	1931	5 Q9NCF9	Q9ncf9 drosophila
20	143	5.7	1931	5 Q9VKH9	Q9vkh9 drosophila
21	142.5	5.6	533	6 Q95JY2	Q95jy2 macaca fasc
22	140.5	5.6	2166	16 Q51465	Q51465 borrelia bu
23	140	5.5	1132	4 Q75065	Q75065 homo sapien
24	140	5.5	1305	10 Q9FJ35	Q9fj35 arabidopsis
25	138.5	5.5	1391	11 Q922J3	Q922j3 mus musculu
26	138	5.5	1084	16 Q83423	Q83423 treponema p
27	138	5.5	1955	5 Q61308	Q61308 parascaris
28	136.5	5.4	684	3 Q07238	Q07238 pneumocysti
29	136.5	5.4	709	5 Q9GRG1	Q9grg1 tetrahymena
30	136	5.4	1388	11 P70336	P70336 mus musculu
31	135.5	5.4	1956	5 Q20641	Q20641 caenorhabdi
32	135	5.3	1133	5 Q21022	Q21022 caenorhabdi
33	135	5.3	1330	6 Q97961	Q97961 vulpes vulp
34	135	5.3	1530	4 Q43241	Q43241 homo sapien
35	135	5.3	5458	5 Q9U459	Q9u459 plasmodium
36	134.5	5.3	1300	4 Q13999	Q13999 homo sapien
37	134.5	5.3	2035	11 Q9JMH9	Q9jmh9 mus musculu
38	134	5.3	1538	13 Q98940	Q98940 gallus gall
39	133.5	5.3	720	4 Q9H6Q7	Q9h6q7 homo sapien
40	133.5	5.3	1410	4 Q14221	Q14221 homo sapien
41	133.5	5.3	2138	5 Q9XZE3	Q9xze3 amoeba prot
42	133	5.3	1938	6 Q9GJP9	Q9gjp9 oryctolagus
43	132.5	5.2	1708	5 Q9U0S6	Q9u0s6 mytilus gal
44	132.5	5.2	2473	11 Q9QZ84	Q9qz84 mus musculu
45	132	5.2	1033	4 Q9UFE1	Q9ufel homo sapien

ALIGNMENTS

RESULT 1

Q15846 PRELIMINARY; PRT; 466 AA.
ID Q15846
AC Q15846; TREMBLrel. 01, Created)
DC 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 54.2 KDA PROTEIN PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NEURETINA;
RA Shimizu A., Nishida K., Kinoshita S., Inazawa J., Okubo K.,
RA Matsubara K.;
RT "Expression profile of active genes in human retina.";
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; D63813; BAA09882.1; -;
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 3.
DR SMART; SM00035; Cla; 1.
DR SMART; SM00030; CLB; 1.
KW Signal; Hypothetical protein.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 466. ROD PHOTORECEPTOR PROTEIN.
SQ SEQUENCE 466 AA; 54214 MW; 41A603D0FD6439C4 CRC64;

Query Match 97.3%; Score 2460; DB 4; Length 466;
Best Local Similarity 99.8%; Pred. No. 2.2e-160;
Matches 465; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 MKPPLLVFVCLLWLKDSHCAPTWKDKTALSENKLSFSEVGEIDAEVVKALTGKQMK 71

DB 1 MKPPLLVFVCLLWLKDSHCAPTWKDKTALSENKLSFSEVGEIDAEVVKALTGKQMK 60

QY 72 IMMEREKEHTNLMSTLKKCRKQKALKLLNEVQHELEERLCRESLADSGECRSCL 131

Db	61	IMMERKEHTNLASTLAKCREEQEALKLLNEVQEHLEBEERLCRESLADSWGECRSL	120
Qy	132	ENNCMRIYTCOPSSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLIKDAQLTQM	191
Db	121	ENNCMRIYTCOPSSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLIKDAQLTQM	180
Qy	192	EDVFSQLTVDVNSLFRNSFNFRMQQEFQDTFQSHFISDLDLTPYFFPAFSKEPMTKA	251
Db	181	EDVFSQLTVDVNSLFRNSFNFRMQQEFQDTFQSHFISDLDLTPYFFPAFSKEPMTKA	240
Qy	252	DLEQCWDIPNFFQFLCNFSVSIYVSSTITKMLKAIEDLPKQKAPDGHGGLISKMLPGQ	311
Db	241	DLEQCWDIPNFFQFLCNFSVSIYVSSTITKMLKAIEDLPKQKAPDGHGGLISKMLPGQ	300
Qy	312	DRGLCGELDQNLSCRCFKFHEKQCQAHLSDEDCPDVPALHTEDEAIRLVNVSNOQYGOI	371
Db	301	DRGLCGELDQNLSCRCFKFHEKQCQAHLSDEDCPDVPALHTEDEAIRLVNVSNOQYGOI	360
Qy	372	LQMTKRKLEDYALVLEKMRGQFGWSELANOAPETEIIFNSIQVVPRIHEGNIKQDETM	431
Db	361	LQMTKRKLEDYALVLEKMRGQFGWSELANOAPETEIIFNSIQVVPRIHEGNIKQDETM	420
Qy	432	MTDLSILPSSNFTLKIPLEESAESSNFTGYVYVAKALQHFKHEKFTW	477
Db	421	MTDLSILPSSNFTLKIPLEESAESSNFTGYVYVAKALQHFKHEKFTW	466
RESULT 2			
ID	Q95KN1	PRELIMINARY; PRT; 465 AA.	
AC	Q95KN1		
DT	01-DEC-2001	(TrEMBLrel. 19, Created)	
DT	01-DEC-2001	(TrEMBLrel. 19, Last sequence update)	
DE	01-DEC-2001	(TrEMBLrel. 19, Last annotation update)	
DE	RETINAL CLUSTERIN-LIKE PROTEIN C11L1B SPLICE VARIANT.		
RL	Gene 243:151-160(2000).		
OC	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_TaxID=9615;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=RETINA;		
RX	MEDLINE=20156379; PubMed=10675623;		
RA	Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Aguirre G.D.;		
RA	"Molecular cloning, characterization and expression of a novel retinal		
RT	clusterin-like protein cDNA.";		
RL	Gene 243:151-160(2000).		
DR	EMBL; AF241221; AAK49030.1; -.		
SQ	SEQUENCE 465 AA; 54402 MW; BBDEIAC512D5D33F CRC64;		
Query Match 72.5%; Score 1832.5; DB 6; Length 465;			
Best Local Similarity 73.2%; Pred. No. 1.7e-117;			
Matches 341; Conservative 52; Mismatches 72; Indels 1; Gaps 1;			
Qy	12	MKPPLLVFVCLLWKDCHCAPTKKDTATSENLSKSEVEIDAADREVKKALTGKQMK	71
Db	1	MKPPLLVFVCLLWKDCHCAPTKKDTATSENLSKSEVEIDAADREVKKALTGKQMK	60
Qy	72	IMMERKEHTNLASTLAKCREEQEALKLLNEVQEHLEBEERLCRESLADSWGECRSL	131
Db	61	IMMERKEHTNLASTLAKCREEQEALKLLNEVQEHLEBEERLCRESLADSWGECRSL	120
Qy	132	ENNCMRIYTCOPSSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLIKDAQLTQM	191
Db	121	ENNCMRIYTCOPSSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLIKDAQLTQM	180
Qy	192	EDVFSQLTVDVNSLFRNSFNFRMQQEFQDTFQSHFISDLDLTPYFFPAFSKEPMTKA	251
Db	181	EDVFSQLTVDVNSLFRNSFNFRMQQEFQDTFQSHFISDLDLTPYFFPAFSKEPMTKA	240
Qy	252	DLEQCWDIPNFFQFLCNFSVSIYVSSTITKMLKAIEDLPKQKAPDGHGGLISKMLPGQ	311

Db	241	DPVQSWDIPSEFFQFYNFSSUIYHSISTITTKTLNAIEDLPKQDNDNHSLSKTLPVQ	300
Qy	312	DRGLCGELDQNLSCRCFKFHEKQCQAHLSDEDCPDVPALHTEDEAIRLVNVSNOQYGOI	371
Db	301	HRGYPGEGQNLSECFQFHARQKQCDYLWEDCPDELHTKVDLEALVNIHQQAQV	360
Qy	372	LQMTKRKLEDYALVLEKMRGQFGWSELANOAPETEIIFNSIQVVPRIHEGNIKQDETM	431
Db	361	LQMTQHLEDITYLMEKMRREEFGWADLANQAPGAENIFDSTKMPNIHEGNEFSKQDETM	420
Qy	432	MTDLSILPSSNFTLKIPLEESAESSNFTGYVYVAKALQHFKHEKFTW	477
Db	421	I-DLSILSSNFTLKIPLEESAETSNTFISYMLEKAVQHFKHEKFTW	465
RESULT 3			
ID	Q9N1T8	PRELIMINARY; PRT; 338 AA.	
AC	Q9N1T8		
DT	01-OCT-2000	(TrEMBLrel. 15, Created)	
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)	
DT	01-JUN-2001	(TrEMBLrel. 17, Last annotation update)	
DE	RETINAL-SPECIFIC CLUSTERIN-LIKE PREPROTEIN.		
GN	C10L1.		
OS	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_TaxID=9615;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20156379; PubMed=10675623;		
RA	Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Aguirre G.D.;		
RA	"Molecular cloning, characterization and expression of a novel retinal		
RT	clusterin-like protein cDNA.";		
RL	Gene 243:151-160(2000).		
DR	EMBL; AF147784; AAF36799.1; -.		
DR	InterPro; IPR000753; Clusterin.		
DR	SMART; SM00035; CLA; 1.		
DR	SMART; SM00030; CLB; 1.		
SQ	SEQUENCE 338 AA; 39385 MW; B86F07877213FCC2 CRC64;		
Query Match 48.5%; Score 1226.5; DB 6; Length 338;			
Best Local Similarity 69.8%; Pred. No. 3e-76;			
Matches 231; Conservative 42; Mismatches 57; Indels 1; Gaps 1;			
Qy	147	SSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLIKDAQLTQMEDVFSQLTVDVNSLF	206
Db	9	TSGSTVEQEFERNIYQYLFPPFEDNEKDLPGVEKFIEDAQVAQIENVFQLTVDVNRFLF	68
Qy	207	NRSFNVRMQQEPDQTFQSHFISDLDLTPYFFPAFSKEPMTKADLEQCWDIPNFFQLF	266
Db	69	NRSFNVRMQQEPDQTFQSHFISDLDLTPYFFPAFSKEPMTKADLEQCWDIPNFFQLF	128
Qy	267	CNFSVSIYVSSTITKMLKAIEDLPKQKAPDGHGGLISKMLPGQDRGLCGELDQNLSC	326
Db	129	YNFSLSYHSISTITTKTLNAIEDLPKQDNDNHSLSKTLPVQHRGYPGEGQNLSEC	188
Qy	327	FKFHEKQCQAHLSDEDCPDVPALHTEDEAIRLVNVSNOQYGOILQMTKRKLEDYALV	386
Db	189	FQFHARQKQCDYLWEDCPDELHTKVDLEALVNIHQQAQVLOHTQHHLEDITYLM	248
Qy	387	EKMRGQFGWSELANOAPETEIIFNSIQVVPRIHEGNIKQDETMMDLSILPSSNFTLK	446
Db	249	EKMRGQFGWADLANQAPGAENIFDSTKMPNIHEGNEFSKQDETM-DLSILSSPNFTLK	307
Qy	447	IPLEESAESSNFTGYVYVAKALQHFKHEKFTW	477
Db	308	IPLEESAETSNTFISYMLEKAVQHFKHEKFTW	338
RESULT 4			

```

Q29482
ID Q29482 PRELIMINARY; PRT; 449 AA.
AC Q29482;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CLUSTERIN PRECURSOR.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OLIVER; TISSUE-TESTIS;
RA Barber J.A., Farris J.A., Troedsson M.H.T., Crabo B.G., Foster D.;
RT "Nucleotide sequence of the complementary DNA encoding Equine
  Clusterin";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR EMBL: L46797; AAA80313.1; -.
DR InterPro: IPR000753; Clusterin.
DR Pfam: PF01093; Clusterin; 1.
DR SMART: SM00035; Cla; 1.
DR SMART: SM00030; Clb; 1.
DR PROSITE: PS00492; CLUSTERIN_1; 1.
DR PROSITE: PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Signal.
FT SIGNAL 1
FT CHAIN 22 449
FT POTENTIAL.
SQ SEQUENCE 449 AA; 52154 MW; 705A60504515F7C0 CRC64;

```

Query Match	16.78;	Score 423;	DB 6;	Length 449;
Best Local Similarity	24.5%;	Pred. No. 3.5e-21;		
Matches 117;	Conservative 100;	Mismatches 203;	Indels 58;	Gaps 13;
Qy	16	LLVFIVCLLMKDSHCAPTWKDKTAISENLKSFSEVGEIDAEVKKKALTGIKQKIMME	75	
		: : :		
Db	4	LLLGVGLLLTLENGQ---VLGDKAVSDRELQEMSTQGSNYINKEIKNALKGVKQIKNIE	60	
		: : :		
Qy	76	RKEKHTNMLMTKKKREEKQELAKLLNEVEHLEEEERLCRESIADSGWGRCSLENNC	135	
		: : : : : :		
Db	61	QTNERSLLGLTLEAKKKKGGALNDTKDSMKLKESQGVCEWTALWECKPCLCOTC	120	
		: : : : : :		
Qy	136	MRIYVTCQPSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKITEKDAQLTQMEDV	194	
		: : : : : :		
Db	121	MKFYARVCRSGSLVGHQLEEFLLNQSPFYFWINGDR-----IDSLLENDRQQTHVLDV	174	
		: : : : : :		
Qy	195	FSQILTVDVNSLNFNSFVFMQOQEFQTFQSHIFISDTDLTEPYFFPAFSEKPMTKADL-	253	
		: : : : : :		
Db	175	-----MQDSFDRASSI-----MDLFDQDRFFT-REPQDTYYSPFS-SPHRRSLL	218	
		: : : : : :		
Qy	254	-----EQCWDIP-----NFFOLFCNFSVSYESVSETITKMLKALIEDLPKQD--KAP	298	
		: : : : : :		
Db	219	FNPKSRFARNIMHPMTVTHLNDMFQPTFMIHQAOQAMNLHLRULDQJLPMTEFSEGD	278	
		: : : : : :		
Qy	299	DHGGLISKMLPGDRLGCGELDQNLSCFKPHEKQCQQAHLSEDC----PDVPALHPTEL	354	
		: : : : : :		
Db	279	NH-----DRTVCKEIRHNSTGCLKMKDQCEKQEIILSDVCSTNNFSQMLROEL	327	
		: : : : : :		
Qy	355	DEATRLVNSVNOQYGOILQMTKRKLEDTAYLVEKMRGQGVGWSLQANQAPETEIFNSIQ	414	
		: : : : : :		
Db	328	NNSLQLAKEFTKLYDELLQSYQEKMLNTSLLKQNLQEOFSWVSOLANLTQGEDQYILQVT	387	
		: : : : : :		
Qy	415	VVPRIHEGNIKQDETMMTDLISILPSSNFTLKIPLSEASNSNIGYVAKALQHFKE	472	
		: : : : : :		
Db	388	TVSS--HNSDSEVPGLTRVVYWKLFDSYPITVTPEVYSRNNPKPMETVAEKLQYRQ	444	
		: : : : : :		

```

RESULT      5
Q9YGP0
ID          Q9YGP0          PRELIMINARY;      PRT;      448 AA.
AC          Q9YGP0;
DT          01-MAY-1999 (TREMBLrel. 10, Created)
DT          01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT          01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE          CLUSTERIN.
OS          Gallus gallus (Chicken).
OC          CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC          OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC          OC Gallus.
OX          NCBI_TaxID=9031;
RN          [1]
          SEQUENCE FROM N. A.
RX          MEDLINE=99134326; PubMed=9933595;
RA          Mahon M.G., Lindstedt K.A., Hermann M., Nimpf J., Schneider W.J.;
RT          "Multiple involvement of clusterin in chicken ovarian follicle
RT          development. Binding to two oocyte-specific members of the low density
RT          lipoprotein receptor gene family.";
RL          J. Biol. Chem. 274:4036-4044(1999).
CC          -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC          OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC          AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC          CELL DEATH (BY SIMILARITY).
CC          -1- SUBUNIT: ANTI-PARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC          SIMILARITY).
CC          -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR          EMBL; AF119370; A017257.1; -.
DR          InterPro; IPR000753; Clusterin.
DR          Pfam; PF01093; Clusterin; 1.
DR          SMART; SM00035; CLA; 1.
DR          SMART; SM00030; CLB; 1.
DR          PROSITE; PS00492; CLUSTERIN_1; 1.
DR          PROSITE; PS00493; CLUSTERIN_2; 1.
DR          GQycoprotecin.
KW          Q9YGP0
SQ          SEQUENCE      448 AA; 51348 MW; 9982873DCA13D27C CRC64;

```

[illegible]

DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE CLUSTERIN (FRAGMENT)
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE DAWLEY; TISSUE=THYMUS;
 RA Park J.H., Park J.S., Ju S.K., Na S.Y., You K.H.;
 RT "Determination of clusterin mRNA expression of apoptosis induced rat
 thymocytes in vivo and in vitro."
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
 OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
 AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
 CELL DEATH (BY SIMILARITY).
 CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
 SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
 DR EMBL; AF314657; AAG31162.1; -
 DR InterPro: IPR000753; Clusterin.
 DR Pfam; PF01093; Clusterin; 2.
 DR SMART; SM00035; Cla; 1.
 DR PROSITE; PS00493; CLUSTERIN_2; 1.
 KW Glycoprotein.
 FT NON_TER
 SQ SEQUENCE 218 AA; 24931 MW; ED2FC4425A510589 CRC64;

 Query Match 8.1%; Score 204; DB 11; Length 218;
 Best Local Similarity 28.3%; Pred. No. 1.4e-06;
 Matches 47; Conservative 36; Mismatches 73; Indels 10; Gaps 3;

 QY 312 DRGLGELDONLSRCFKFHFHCKQKQAHLSDEC-----PDVPAHLTELDEAIRLVNYSNQ 367
 Db 51 DRTVCBEIRHNTSGCLKMKGQCEKILSDGCTNNPAQANRLNDSLQVAERLTQ 110
 QY 368 YGQILQMTKHEDTAYLVKMKRGQGWSELAN--QAPETELFIENSQVVPRIHESNIS 425
 Db 111 YNELLSLOSKMLNTSLLQLNDQFSWSQLANLQGGDQYLRVSTVT-----HSSDSE 166
 QY 426 KQDETMMTDLUSILPSSNFTLKIPLESASSNFYGVVAKALQHFK 471
 Db 167 VPSRVTEVVVKLSDPITVVLPEEVSQKPKFMDTVAEKALQYR 212

 RESULT 9
 ID 076329 PRELIMINARY; PRT; 1738 AA.
 AC 076329;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE INTERAPTIN.
 GN ABPD.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=98365468; PubMed=9700162;
 RX Rivero F.J., Kuspa A., Brokamp R., Matzner M., Noegel A.A.;
 RT "Interaptin, an actin-binding protein of the alpha-actinin superfamily
 RT in Dictyostelium discoideum, is developmentally and camp-regulated and
 RT associates with intracellular membrane compartments.";
 RL J. Cell Biol. 142:733-750(1998).
 DR EMBL; AF057019; AAC34582.1; -
 DR HSSP; Q01082; 1BKR.
 DR InterPro; IPR001589; Actinin_act_bind.
 DR InterPro; IPR001715; Calponin_hom.

DR InterPro: IPR001990; Granin.
 DR InterPro: IPR001451; Hexapep_transf.
 DR InterPro: IPR002017; Spectrin.
 DR Pfam; PF00307; CH; 2.
 DR SMART; SM00033; CH; 2.
 DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
 DR PROSITE; PS00020; ACTININ_2; 1.
 DR PROSITE; PS00021; CH; 2.
 DR PROSITE; PS00422; GRANINS_1; UNKNOWN_1.
 DR PROSITE; PS00101; HEXAPEP_TRANSFERRASES; UNKNOWN_1.
 SQ SEQUENCE 1738 AA; 204427 MW; 577A99D2EC79AF5C CRC64;

 Query Match 6.5%; Score 164; DB 5; Length 1738;
 Best Local Similarity 21.1%; Pred. No. 0.0087;
 Matches 103; Conservative 87; Mismatches 171; Indels 128; Gaps 21;

 QY 36 KDKTAISENL-----KSFSEVGEIDADEEYKALTGIKOMKIMMER----- 76
 Db 667 KDNQTINEQLNQLSEKDEIKLSNQEQOQDEKINNLLLEKEKDCILIERINOQLLEN 726
 QY 77 -----KEKHTNLMSTLKKREEKQBAKLLNVEQHELEER 114
 Db 727 IDLSKYQQLLLEFENFKLNSKEK-----NOLNELQSKQDERFNQLND--EKLEKEQ 779
 QY 115 LCRESLADSWGECRSCLENNCRITYTTCPSWSVKNKIERFRKIYQFLFPFHEDNEKD 174
 Db 780 L--QSIEDENFYK-----QQQLSSNSN-IDQQQLQSTIIELSELKEQKELN 822
 QY 175 LPISEKLIKDAQLTQMEDVFSOLT-----VDVNSLFNRSFNFMQOQEFQDTFQSH 227
 Db 823 ---DSLIEKEKQLQLOQEFQDLNEKNKQKHDQDLELLEKQ---LKQLQEQYDQLNETN 876
 QY 228 FTSDDTLTPYFPFAPSKPMTKADLE-----QCWDIPNFQFCNFVSIVYSV 277
 Db 877 QSIENOLNQNL---INKENLKEQEELKQLNQLNQIEKIQFDQOEFQSKNSINIELV 933
 QY 278 SETITKMLKAIED---LPKQDKA---PDHGGILSKMLPGQDRGLCGELDONLSRCFKPHE 331
 Db 934 NEKNEKLIQLOQDYDQLKQKQNSNDEKENDLIEK--ENQLKSINQLNOLIEKNESDHK 991
 QY 332 CKCKQQAHLSEDCPD---VPALHTELDE---AIRLVNVSNQYQGIL-----Q 373
 Db 992 EQQLKQSQSTENDLIEKENIQIQLSQNLNEQKQNSQLSEKQQLNQLIEKQFDQKEQ 1051
 QY 374 MTRKHLEDYAY---LVEKMRQFGF-WYSELANQAPETELFIENSQVVPRIHESNISQD 428
 Db 1052 LKQSQSTENDLFEKENIQIQLSQNLNEQKQNSQLSEK-----QQLNOLIEKNESDQK 1105

 QY 429 ETMTMTDLSI 437
 Db 1106 EQQLKQSQSI 1114

 RESULT 10
 ID 09W6V0 PRELIMINARY; PRT; 3616 AA.
 AC 09W6V0;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE HYPERION PROTEIN, 419 KDA ISOFORM.
 GN HYPERION.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=RETINA;
 RA Kemmer W.A., Schwarz U.;
 RT "Characterization of Hyperion, a gene coding for an abundance of gene

DR InterPro: IPR0013439; ABC transportr.
DR InterPro: IPR001687; ATP_GTP_A.
DR InterPro: IPR003716; RNA_Pol_Omega.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR InterPro: IPR002017; Spectrin.
DR Pfam: PF02483; SMC_C; 1.
DR Pfam: PF02463; SMC_N; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1156 AA; 135563 MW; B12DB30F70C0CE49 CRC64;

[illegible]

SQ SEQUENCE 840 AA; 97587 MW; A875627BE3D4D716 CRC64;

Query Match 5.9%; Score 149.5; DB 6; Length 840;

Best Local Similarity 18.6%; Pred. No. 0.036; Mismatches 154; Indels 143; Gaps 18;

Matches 88; Conservative 89; Mismatches 154; Indels 143; Gaps 18;

QY 39 TAISENLKSFSEVGEIDADEEVKALGTGKQKIMMERKEKHTNLMSTLKKCRE-EKQE 97

DB 426 TEMQKNKSVSY--LEMDKTLKKEEVKRLQQLRKQEKVYASALDILLKREKETQEQE 483

QY 98 ALKLLNEVQE---HLEERERLCRESLADSWGECRSCLENNCMRIYTCQPSWSSVKNKI 153

DB 484 FLSLQEEFQKRDKANLEERQKL-----KSRLL 509

QY 154 ERFFRKIYFLPFPHEDNEKDLPISEKLEKDAQLTOMEDVFSQLTVDVNSLNRSE--- 210

DB 510 EKLTLQVKNLQPMSENERAKNLKQO-----QINEVKNKKLQHVARSSEQNVYVK 562

QY 211 NVFRMQQEFQDTFQSHFISDTLTPYFPFAPS---REPMTKADLE----- 254

DB 563 SETAQLKEQLEEVKMSDITKTKMTHSNLLDSCPEERESLAPADIERSSQLASKMHSLL 622

QY 255 -----QCWDIPNPFQFCNFSVSIYESVETITKMLKAI-----EDLPKQDRAP 298

DB 623 ALMVGILLKQODITNSDAE--HFKESS--EKVSDIMLQRLKSLHKKKLLDKKLLKHDRIT 678

QY 299 DHGGLSKMLPGDRLG-----CGE----- 318

DB 679 TFRDILIAKEKAFQDHAUKVTDGDSDEAKSIRDVPTFLGAKLKYHSLNEDELFLITKLGC 738

QY 319 -LDONLSRCFKFHEKQKQAHLSDECDVPALHTDELDAIRLVNVSNOQYQOILQMTKR 377

DB 739 LLESKESHCRNLEENDKYQRLGSLIKKVTSEETIECADQRLAISHSOIAH-LEKRNK 797

QY 378 HLEDYALIVEKMRGQGWVSELANQAPETEIFN---SIQVPRHEGNLSKQD 428

DB 798 HLED---LIRKPR-----EKARK-PRSKSLENHPKSMTMPAVFKNRNDLD 840

RESULT 13

Q28021 PRELIMINARY; PRT: 1388 AA.

ID Q28021

AC Q28021

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE RHO-ASSOCIATED KINASE.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;

OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;

RN [1]

SEQUENCE FROM N.A.

RP MEDLINE=96208507; PubMed=8641286;

RA Matsui T., Amano M., Yamamoto T., Chihara K., Nakafuku M., Ito M.,

RA Nakano T., Okawa K., Iwamatsu A., Kaibuchi K.;

RT "Rho-associated kinase, a novel serine/threonine kinase, as a putative

target for small GTP binding protein Rho.";

RL EMBO J. 15:2208-2216(1996).

CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

DR EMBL; U36909; AAC48567.1; -

DR HSSP; Q63450; 1A06.

DR InterPro; IPR002219; DAG_PE-bind.

DR InterPro; IPR000719; Euk_pkinase.

DR InterPro; IPR001849; PH.

DR InterPro; IPR000961; Pkinase_C.

DR InterPro; IPR000861; REM_repeat.

DR InterPro; IPR002290; Ser_thr_pkinase.

DR Pfam; PF02185; HRI.1.

DR Pfam; PF00169; PH.1.

pfam; PF00069; pkinase; 1.
DR SMART; SM00109; C1; 1.
DR SMART; SM00074; HRI; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00220; S-TKC; 1.
DR SMART; SM00133; S-TKC; 1.
DR PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1388 AA; 160799 MW; CAB3CE7D3860465D CRC64;

Query Match 5.9%; Score 149; DB 6; Length 1388;

Best Local Similarity 16.7%; Pred. No. 0.071;

Matches 95; Conservative 96; Mismatches 166; Indels 212; Gaps 20;

QY 36 KDKTAISENLKSFSEVGEIDADEEVKALGTGKQKMI-----MMERKEKEHTNLS 86

DB 780 KOKDVLNEDVRNLT---LKTEQETQKRLTQNDLKMTQVNTLMKSEKQLKQENHLL 835

QY 87 TLK-----KCREKQKALKLNEVQHELEEE-----RLCR 117

DB 836 EMKMSLEKQNAELRRQDADQMKELQDLEAEQYFSTLYKTVRELKEECEKTKLCK 895

QY 118 E-----SLADSWGECRSCLENNCMRIYTCQPSWSSVKNKLERFERKIYQFLFPFHEDN 171

DB 896 ELQKKQKQELQDE-----RDSLAQLEITLTKADSEQLARSTAEQYSDLEKE 942

QY 172 E--KDLPISE-----KLIEXDAQLTQMEDVFSQLTVDVNSLFRNPNVFRMQQEFQ 222

DB 943 KIMKELEIKEMMARHKQELTEKDATIASLEETNRTLTSVANLANEKEELNNKLKEAQO 1002

QY 223 TQSHFISDTLTPYFPFAPSKEPMTKADLEQCWDIPNFFQFCNFSVSIYESVETIT 282

DB 1003 LSR---LKDEEISAAAIKQAFKQLLTERTLK-----TQAVN 1036

QY 283 KMLKAIE-----DLPKQKAPDHGGLSKMLPGQDRGLGCEL---DONLSRCF 327

DB 1037 KLAETMNRKEPVKRGNDTDRRKE-----ENRKLHMLKSEREKLQMI 1082

QY 328 KPEHKCQKQAHLSDECDVPALHTDELDAIRLVNVSNOQYQOILQMTKRHLE----- 380

DB 1083 KYOKELNEMQAIAEE---SOIRTELQMTLDSKDSIDIEQLRSQALHIGLDSSSISGS 1138

QY 381 -----DTAYLIVEKMRG-----QFGWVSEL----- 399

DB 1139 PGDTEADGFPESRLEGWLSLVRNNTKFGWVKYIVVSSKKILFYDSEQDKESNPYM 1198

QY 400 -----ANQAPETEIFNSIQVVPRI-----HEGNISKQDETMMTDLSILPSSNFT 444

DB 1199 VLDIDKLFHVRPVTQTDVYRADAKEIPRFQILYANEGESKKEQE----- 1243

QY 445 LKIPLEESAESSNFI---GYVVAKALQHF 470

DB 1244 --FPVPEVGEKSNYICHKGHEFIPTLYHF 1270

RESULT 14

Q18082 PRELIMINARY; PRT: 1091 AA.

ID Q18082

AC Q18082;

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE HYPOTHETICAL 127.0 KDA PROTEIN.

GN C18C4.5.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:32:26 ; Search time 56.55 seconds
(without alignments)
972.264 Million cell updates/sec

Title: US-09-722-544A-2
Perfect score: 2632
Sequence: 1 MKIKAEKNEGSPRSWQLHW.....FIGYVAKALQHFKEHFKTW 495

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneq/geneq-emb/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2586	98.3	495	AAV30785	Protein encoded by
2	2464	93.6	466	AAV30793	Immature human HKN
3	2447	93.0	477	AAV30786	Protein encoded by
4	2350	89.3	446	AAV30792	Mature secreted hum
5	1757.5	66.8	465	AAV30791	Bovine HKNG1 ortho
6	1639	62.3	466	AAV30787	Protein encoded by
7	1558	59.2	450	AAV30788	Guinea pig HKNG1 o
8	1277	48.5	374	AAV30789	Guinea pig HKNG1 o
9	1272.5	48.3	373	AAV30790	Guinea pig HKNG1 o
10	925.5	35.2	521	AAV30794	Amino acid sequenc
11	434	16.5	449	AAV28048	Novel human secret

12	432	16.4	448	12	AAV11704	Cytolysis Inhibito
13	405.5	15.4	416	22	ABBS0285	Apolipoprotein J o
14	400.5	15.2	446	21	AA803441	Porcine clusterin
15	320	12.2	363	22	AAE03764	Human gene 1 encod
16	173.5	6.6	148	22	AAO00607	Human polypeptide
17	160.5	6.1	116	21	AAO30745	Human secreted pro
18	159.5	6.1	139	22	AAE03783	Human gene 1 encod
19	159.5	6.1	139	22	AAE01703	Human gene 4 encod
20	152.5	5.8	247	22	AAE03787	Human gene 1 encod
21	145	5.5	1372	19	AAW56473	Protein with Rho p
22	143	5.4	1931	22	ABB61012	Drosophila melanog
23	142.5	5.4	2633	22	ABG06505	Novel human diagno
24	141.5	5.4	3433	18	AAW22017	utrophin. Homo sa
25	141	5.4	944	21	AAV67600	Human adipose tiss
26	141	5.4	1388	19	AAW56475	Protein with Rho p
27	139.5	5.3	2663	22	AAW39097	Human polypeptide
28	139.5	5.3	2688	22	AAW40883	Human polypeptide
29	139	5.3	1374	22	AAW69070	Human male enhance
30	136	5.2	934	22	AAU01768	Human secreted pro
31	136	5.2	2517	21	AAV71159	Human phosphodiast
32	135.5	5.1	1427	12	AAV10534	Human 160kb mediat
33	134	5.1	5373	22	AAU14603	Novel bone marrow
34	134	5.1	5447	22	AAU14697	Human polypeptide
35	130.5	5.0	1788	22	AAW40467	Human mitotin. Ho
36	130.5	5.0	2482	16	AAW72826	Human mitotin amin
37	130.5	5.0	2482	19	AAW23996	Human protein SEQ
38	130	4.9	990	22	AAW78520	AMML chromosome in
39	129	4.9	816	16	AAW66931	Restin protein seq
40	128.5	4.9	1392	20	AAW06999	Nucleolar/endosoma
41	128.5	4.9	1411	17	AAW02258	Kinetochore protei
42	128.5	4.9	3248	17	AAW99795	AMML chromosome in
43	128	4.9	885	16	AAW66930	Cellular homologue
44	127	4.8	1047	18	AAW01535	Human polypeptide
45	126.5	4.8	1780	22	AAW38681	

ALIGNMENTS

RESULT 1
AAV30785
ID AAV30785 standard; Protein; 495 AA.
XX
AC AAV30785;
XX
DT 23-NOV-1999 (first entry)
XX
DE Protein encoded by human HKNG1 cDNA.
DE
KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW schizophrenia; splice variant.
XX
OS Homo sapiens.
XX
PN WO9947535-A1.
XX
PD 23-SEP-1999.
XX
PF 16-MAR-1999; 99WO-US05606.
XX
PR 16-MAR-1998; 98US-0078044.
PR 05-JUN-1998; 98US-0088312.
PR 28-OCT-1998; 98US-0106056.
PR 22-JAN-1999; 99US-0236134.
XX
(MILL-) MILLENNIUM PHARM INC.
(REGC) UNIV CALIFORNIA.
PA
PI Chen H, Freimer NB;
XX
DR WPI; 1999-562047/47.
DR N-PSDB; AA210750.

XX	PT	New HKNG1 polynucleotides useful in diagnosis and treatment of
XX	PT	neuropsychiatric disorders, e.g. bipolar affective disorders and
XX	PT	schizophrenia
XX	PS	Claim 1; Fig 1A-B; 205pp; English.
XX	CC	The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1
XX	CC	is a gene associated with bipolar affective disorder (BAD). HKNG1
XX	CC	polynucleotides are useful to identify compounds modulating HKNG1
XX	CC	expression or HKNG1 polypeptide expression/activity. Compounds inhibiting
XX	CC	or enhancing HKNG1 gene expression or activity in individuals can then
XX	CC	be administered therapeutically to treat HKNG1-mediated disorders,
XX	CC	especially neuropsychiatric disorders e.g. BAD, schizophrenia, or
XX	CC	HKNG1-mediated myopia disorders, such as early-onset autosomal
XX	CC	dominant myopia. The polynucleotides can be used in gene therapy
XX	CC	techniques to treat such disorders. They are also useful in diagnosis
XX	CC	to identify individuals having, or at risk of developing, HKNG1-mediated
XX	CC	disorders due to mutations in the HKNG1 gene. Such mutations especially
XX	CC	result in the production of a protein with a different sequence to
XX	CC	the human full-length HKNG1 polypeptide or splice variant sequences,
XX	CC	especially the substitution of a lysine for a glutamic acid at residue
XX	CC	202 or 184. The polynucleotides are also useful in gene mapping, to
XX	CC	produce probes or primers to identify similar sequences (e.g. mutants
XX	CC	or sequences from different species) and to produce transgenic
XX	XX	animals.
XX	SQ	Sequence 495 AA;
		Query Match 98.3%; Score 2586; DB 20; Length 495;
		Best Local Similarity 99.2%; Pred. No. 1.6e-205;
		Matches 490; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY	1	MKTKAEKNEGPRSWQLHWGDIANNNGMKPPLLVFIVCLLWLDKSHCAPTWKDKTATS 60
Db	1	mkikaeknegprswqlhwgdiannngmkppllvfivclllwldkshcceptwkdkats 60
QY	61	ENUKSFSEGEIDAEVVKALTGIFKQKIMMERKEKEHTNLMSTLKKREEKQEAALKLL 120
Db	61	enlksfsevgelidadeevkaltgikqkimmerkekehtnlmstlkkreekqealkll 120
QY	121	NEVOEHLEEEERLCRESLADSWGECRSCLENNCMRTYTTCPQSMSSVKNKIERFRKIYQ 180
Db	121	nevqehleeeerlcrsladshgecrscleenncmrlyttcqpssvsvknkierfrkiyq 180
QY	181	FLPFFHEDNEKDLPISEKLTIEDAQLTQMEDVFSQLTVDVNSLNFNRSFNVFRMQQEFQDQ 240
Db	181	flpfhbednekdlipsekleedaqltqkedvfgsltdvnslnfnrsvnfrqkqepdq 240
QY	241	TFQSHFISDTDLTEPYFFPAFSPKPMKADLEQCWDIPNFFQLFCNFSVSIYSESVETIT 300
Db	241	tfqshfisdtdltepyffpafskpmtkadleqcwdipnffqlfcnfsvsiyvesvett 300
QY	301	KMLKATEDLPKQDKAPDHGLISKMLPGODRGICGELDONLSRCFKFHEKCKQCAHLSE 360
Db	301	kmlkatedlpkqdkapdhggllskmlpgdgorgicgeldqnlsrcfkfhekckcqahlse 360
QY	361	DCPDVPALHTELDEAIRLVNSVNSQOYCGILQMTKRKLEDTAYLVKMRGQFGWVSELANQ 420
Db	361	dcpdvpalteldeairlvnsvnsqyqgilqmrkrkledtaylvkrmrgqfgwvselanq 420
QY	421	APTEIFINSIQVVPRIHGNISKQDETMTWTDLSILPSSNFTLKIPLSESAESSNFIGYV 480
Db	421	apetelfinsiqvvprihgnisqkdetmttdlsilpssnftlkipleesaessnfigyv 480
QY	481	VAKALQHEKEHFKT 494
Db	481	vakalqnfkehft 494

RESULT 2
AAY30793

ID	AAY30793 standard; Protein; 466 AA.
XX	
AC	AAY30793;
XX	
DT	23-NOV-1999 (first entry)
XX	
DE	Immature human HKNG1 protein form 2.
XX	
KW	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW	schizophrenia; splice variant.
XX	
OS	Homo sapiens.
XX	
PN	W09947535-A1.
XX	
PD	23-SEP-1999.
XX	
PF	16-MAR-1999; 99WO-US05606.
XX	
PR	16-MAR-1998; 98US-0078044.
PR	05-JUN-1998; 98US-0088312.
PR	28-OCT-1998; 98US-0106056.
PR	22-JAN-1999; 99US-0236134.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
PA	(REGC) UNIV CALIFORNIA.
XX	
PI	Chen H, Freimer NB;
XX	
DR	WPI; 1999-562047/47.
XX	
PT	New HKNG1 polynucleotides useful in diagnosis and treatment of
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and
PT	schizophrenia
XX	
PS	Claim 22; Fig 17; 205pp; English.
XX	
CC	The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1
CC	is a gene associated with bipolar affective disorder (BAD). HKNG1
CC	polynucleotides are useful to identify compounds modulating HKNG1
CC	expression or HKNG1 polypeptide expression/activity. Compounds inhibiting
CC	or enhancing HKNG1 gene expression or activity in individuals can then
CC	be administered therapeutically to treat HKNG1-mediated disorders,
CC	especially neuropsychiatric disorders e.g. BAD, schizophrenia, or
CC	HKNG1-mediated myopia disorders, such as early-onset autosomal
CC	dominant myopia. The polynucleotides can be used in gene therapy
CC	techniques to treat such disorders. They are also useful in diagnosis
CC	to identify individuals having, or at risk of developing, HKNG1-mediated
CC	disorders due to mutations in the HKNG1 gene. Such mutations especially
CC	result in the production of a protein with a different sequence to
CC	the human full-length HKNG1 polypeptide or splice variant sequences,
CC	especially the substitution of a lysine for a glutamic acid at residue
CC	202 or 184. The polynucleotides are also useful in gene mapping, to
CC	produce probes or primers to identify similar sequences (e.g. mutants
CC	or sequences from different species) and to produce transgenic
CC	animals.
XX	
SQ	Sequence 466 AA;
Query Match 93.6%; Score 2464; DB 20; Length 466;	
Best Local Similarity 100.0%; Pred. No. 1.9e-195;	
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	30 MKPPLLVFIVCLLWLDKSHCAPTWKDKTATSENLSKFSFSEVGEIDAEVVKALTGKQMK 89
Db	1 mkppllvfivclllwldkshcceptwkdkatsenlskfsfsevgelidadeevkaltgkqm 60
QY	90 IMMERKEKEHTNLMSTLKKREEKQEAALKLLNEVQHELEEEERLCRESLADSWGECRSC 149
Db	61 immerkekehtnlmstlkkreekqealkllnevqheleeeerlcrsladswgecrscl 120

QY 150 ENNCMIYTTCPSSVSKNIERFRKIYQFLFPFHEDNEKDLPISEKLEEDAQFTQM 209
 Db 121 enncmriyttcpsvsvknierfrkiyqflfpfhednekdipisekleedaqltqm 180
 QY 210 EDVFSQLVDVNSLFRNFVNRQMQEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 269
 Db 181 edvfsqldvnslnfrsfnvrmqgefddtfgshfisdtdltepyffpafskpemtka 240
 QY 270 DLEQCWDIPNFFQFCNFSVSIYESVETITKMLKAIEDLPKQDKAPDHPGGLISKMLPGQ 329
 Db 241 dleqcwdipnffqfcnfsvsiyevsetitkmlkaiedlpkqdkapdhpghglskmlpgq 300
 QY 330 DRGLGELDQNLSCRFKHEKQCQCAHLSEDCPDVPALHTEDEAIRLVNVSQYQGI 389
 Db 301 drglcgeldqnlscrfkhekqcahlsedcpdvpalhhteldeairlvnvsqyqgi 360
 QY 390 LQMRKHLEDYAYLVEKMRGQFGWVSELANQAPETIIFNSIQVVPRIHEGNIKSQDET 449
 Db 361 lqmrkhledtayivekmrgqfgwvselanqapetiiifnsiqvvprihegniskqdet 420
 QY 450 MTDLSILPSSNFTLKIPLEESAESSNFYGYVVAQALQHFKEHFKTW 495
 Db 421 mtdlsilpssnftlkipleesaessnfgyvvakalqhfkehfktx 466

RESULT 3

AAY30786
 ID AAY30786 standard; Protein; 477 AA.

XX AC AAY30786;

DT 23-NOV-1999 (first entry)

DE Protein encoded by a human HKNG1 splice variant HKNG1-V1.

KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
 KW schizophrenia; splice variant.

XX OS Homo sapiens.

PN WO9947535-A1.

PD 23-SEP-1999.

XX 16-MAR-1999; 99WO-US05606.

PR 16-MAR-1998; 98US-0078044.

PR 05-JUN-1998; 98US-0088312.

PR 28-OCT-1998; 98US-0106056.

XX 22-JAN-1999; 99US-0236134.

PA (MILL-) MILLENNIUM PHARM INC.

PA (REGC) UNIV CALIFORNIA.

PI Chen H, Freimer NB;

XX WPI; 1999-562047/47.

DR N-PSDB; AAZ10751.

XX New HKNG1 polynucleotides useful in diagnosis and treatment of

PT neuropsychiatric disorders, e.g. bipolar affective disorders and
 PT schizophrenia -
 XX Claim 1; Fig 2A-B; 205pp; English.
 XX The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)
 CC splice variant. HKNG1 is a gene associated with bipolar affective
 CC disorder (BAD). HKNG1 polynucleotides are useful to identify compounds
 CC modulating HKNG1 gene expression or HKNG1 polypeptide expression/
 CC activity. Compounds inhibiting or enhancing HKNG1 gene expression or
 CC activity in individuals can then be administered therapeutically to
 CC treat HKNG1-mediated disorders, especially neuropsychiatric disorders

CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as
 CC early-onset autosomal dominant myopia. The polynucleotides can be used
 CC in gene therapy techniques to treat such disorders. They are also useful
 CC in diagnosis to identify individuals having, or at risk of developing,
 CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such
 CC mutations especially result in the production of a protein with a
 CC different sequence to the human full-length HKNG1 polypeptide or
 CC splice variant sequences, especially the substitution of a lysine for
 CC a glutamic acid at residue 202 or 184. The polynucleotides are also
 CC useful in gene mapping, to produce probes or primers to identify
 CC similar sequences (e.g. mutants or sequences from different species)
 CC and to produce transgenic animals.

XX Sequence 477 AA;

Query Match 93.0%; Score 2447; DB 20; Length 477;
 Best Local Similarity 99.1%; Pred. No. 4.9e-194;
 Matches 466; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 25 NNSGNMKPPLLVFVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGETDADEEVKKALTG 84
 Db 7 nsngnmkpplllvfvcillwlkdschaptwkdktaisenlksfsevgeidaadeevkkaltg 66
 QY 85 IKQMKIMMERKEKEHTNLMSTLKKCREKQKALKLLNEVQEHLEERLCRESLADSMGE 144
 Db 67 ikqmkimmerkekehtnlmstlkkcreekqaalkllnevehleeeerlcresladswe 126
 QY 145 CRSCLENNCMIYTTCPSSVSKNIERFRKIYQFLFPFHEDNEKDLPISEKLEIEDA 204
 Db 127 crsclenncmriyttcpsvsvknierfrkiyqflfpfhednekdipisekleieeda 186
 QY 205 QLTQMEDVFSOLTVDVNSLFRNFVNRQMQEFDQTFQSHFISDTDLTEPYFFPAFSKE 264
 Db 187 qltqmedvfsqldvnslnfrsfnvrmqgefddtfgshfisdtdltepyffpafske 246
 QY 265 PMTKADLEQCWDIPNFFQFCNFSVSIYESVETITKMLKAIEDLPKQDKAPDHPGGLISK 324
 Db 247 pmtkadleqcwdipnffqfcnfsvsiyevsetitkmlkaiedlpkqdkapdhpghglsk 306
 QY 325 MLPQDRLGCLGELDQNLSCRFKHEKQCQCAHLSEDCPDVPALHTEDEAIRLVNVSQ 384
 Db 307 mlpqdrglclgeldqnlscrfkhekqcahlsedcpdvpalhhteldeairlvnvsq 366
 QY 385 QYGQILQMRKHLEDYAYLVEKMRGQFGWVSELANQAPETIIFNSIQVVPRIHEGNI 444
 Db 367 qygqilqmrkhledtayivekmrgqfgwvselanqapetiiifnsiqvvprihegnisk 426
 QY 445 QDETMMTDLISLPSSNFTLKIPLEESAESSNFYGYVVAQALQHFKEHFKT 494
 Db 427 qdetmmtdlsilpssnftlkipleesaessnfgyvvakalqhfkehfktx 476

RESULT 4

AAY30792
 ID AAY30792 standard; Protein; 446 AA.

XX AC AAY30792;

DT 23-NOV-1999 (first entry)

DE Mature secreted human HKNG1 protein sequence.

KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
 KW schizophrenia; splice variant.

XX OS Homo sapiens.

XX WO9947535-A1.

PD 23-SEP-1999.

PF	16-MAR-1999;	99WO-US05606.	
XX			
PR	16-MAR-1998;	98US-0078044.	
PR	05-JUN-1998;	98US-0088312.	
PR	28-OCT-1998;	98US-0106056.	
PR	22-JAN-1999;	99US-0236134.	
XX			
PA	(MILL-) MILLENNIUM PHARM INC.		
XX	(REGC) UNIV CALIFORNIA.		
PI	Chen H, Freimer NB;		
XX			
DR	WPI; 1999-562047/47.		
XX			
PT	New HKNG1 polynucleotides useful in diagnosis and treatment of		
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and		
PT	schizophrenia -		
XX			
PS	Claim 5; Fig 17; 205pp; English.		
XX			
CC	The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1		
CC	is a gene associated with bipolar affective disorder (BAD). HKNG1		
CC	polynucleotides are useful to identify compounds modulating HKNG1 gene		
CC	expression or HKNG1 polypeptide expression/activity. Compounds inhibiting		
CC	or enhancing HKNG1 gene expression or activity in individuals can then		
CC	be administered therapeutically to treat HKNG1-mediated disorders,		
CC	especially neuropsychiatric disorders e.g. BAD, schizophrenia, or		
CC	HKNG1-mediated myopia disorders, such as early-onset autosomal		
CC	dominant myopia. The polynucleotides can be used in gene therapy		
CC	techniques to treat such disorders. They are also useful in diagnosis		
CC	to identify individuals having, or at risk of developing, HKNG1-mediated		
CC	disorders due to mutations in the HKNG1 gene. Such mutations especially		
CC	result in the production of a protein with a different sequence to		
CC	the human full-length HKNG1 polypeptide or splice variant sequence,		
CC	especially the substitution of a lysine for a glutamic acid at residue		
CC	202 or 184. The polynucleotides are also useful in gene mapping, to		
CC	produce probes or primers to identify similar sequences (e.g. mutants		
CC	or sequences from different species) and to produce transgenic		
CC	animals.		
XX			
SQ	Sequence 446 AA;		
<hr/>			
Query Match 89.3%; Score 2350; DB 20; Length 446;			
Best Local Similarity 100.0%; Pred. No. 4.8e-186;			
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	50	APTWKDKTAISENLKSFSEYGEIDADEVKKALTGKQKIMMERKEKHTNLMSTLKKC	109
DB	1	aptwkdktaisenlksfseveidadeevkkaltgikqkimmerkekehtnlmstlkkc	60
QY	110	REEKQEKALLNEVQEHLEERLCRESLADSWGECRSCLENMRYTTTCQPSWSSVKN	169
DB	61	reekqealkilnevqehleerlcrsladswgecrscleennmrytttcqpswssvkn	120
QY	170	KIEREFKIKYQFPPFHEDNEKDLPISEKLEEDAQLTQMEDVFSQLTVDVNSLFRSFN	229
DB	121	kierfrkiyqfppfhednekdplisekleedaqltqmedvfsqltvdvnslnfrsfn	180
QY	230	VFRMQQEFQDTQSFHSISDITETPEYFFPAFSKEPMTKADLEQCWDIPNFFQIFCNFSV	289
DB	181	vfrmqqefqdtqsfhsisditetyffpafskpmtkadleqcwdipnffqlfcnfsv	240
QY	290	SIYESVSETTKMLKAIEDLPKQKAPDHGGLISKMLPGODRGLCGELDONLSRCFKFHE	349
DB	241	siyevsetitkmlkaiedlpkqkaphdggliskmlpgqdrglcgeldnlsrckfkhe	300
QY	350	KCKQCAHLSEDPCDVPALHTEDEATRLVNSNQVGOILQMTKRHLEDTAYLVEKMRG	409
DB	301	kckqcahlstedcpvpalhteideatrlvnvsnqvgqilqmtkrhleidtaylvekmrg	360
QY	410	QFGWVSELANQAPETEIFNSIQVVPRIHGNISKQDETMMTDLISILPSSNFTLKIPLEE	469
XX			

Db	361	qfgwvselanqapetelifnsiqvvprihgniskqdetmmtdlslpssnftlkiplee	420
QY	470	SAESSNFIGYVVAQALQHFKHEFKTW	495
DB	421	saessnfigyvvakalqhfkefkfw	446
<hr/>			
RESULT	5		
AAAY30791			
ID	AAAY30791	standard; Protein; 465 AA.	
XX			
AC	AAAY30791;		
XX			
DT	23-NOV-1999	(first entry)	
XX			
DE	Bovine HKNG1 ortholog splice variant protein.		
XX			
KW	HKNG1: Hong Kong new gene 1; bipolar affective disorder; BAD;		
KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;		
KW	schizophrenia; splice variant.		
XX			
OS	Bos sp.		
XX			
PN	WO9947535-A1.		
XX			
PD	23-SEP-1999.		
XX			
PF	16-MAR-1999;	99WO-US05606.	
XX			
PR	16-MAR-1998;	98US-0078044.	
PR	05-JUN-1998;	98US-0088312.	
PR	28-OCT-1998;	98US-0106056.	
PR	22-JAN-1999;	99US-0236134.	
XX			
PA	(MILL-) MILLENNIUM PHARM INC.		
PA	(REGC) UNIV CALIFORNIA.		
XX			
PI	Chen H, Freimer NB;		
XX			
DR	WPI; 1999-562047/47.		
DR	N-PSDB; AA210759, AA210760, AA210761.		
XX			
PT	New HKNG1 polynucleotides useful in diagnosis and treatment of		
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and		
PT	schizophrenia -		
XX			
PS	Claim 1; Fig 11A-B; 205pp; English.		
XX			
CC	The present sequence is encoded by bovine HKNG1 ortholog splice		
CC	variant. HKNG1 (Hong Kong new gene 1) is a gene associated		
CC	with bipolar affective disorder (BAD). HKNG1 polynucleotides are		
CC	useful to identify compounds modulating HKNG1 gene expression or		
CC	HKNG1 polypeptide expression/activity. Compounds inhibiting or		
CC	enhancing HKNG1 gene expression or activity in individuals can		
CC	then be administered therapeutically to treat HKNG1-mediated		
CC	disorders, especially neuropsychiatric disorders e.g. BAD,		
CC	schizophrenia, or HKNG1-mediated myopia disorders, such as		
CC	early-onset autosomal dominant myopia. The polynucleotides can be used		
CC	in gene therapy techniques to treat such disorders. They are also useful		
CC	in diagnosis to identify individuals having, or at risk of developing,		
CC	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such		
CC	mutations especially result in the production of a protein with a		
CC	different sequence to the human full-length HKNG1 polypeptide or		
CC	splice variant sequences, especially the substitution of a lysine for		
CC	a glutamic acid at residue 202 or 184. The polynucleotides are also		
CC	useful in gene mapping, to produce probes or primers to identify		
CC	similar sequences (e.g. mutants or sequences from different species)		
CC	and to produce transgenic animals.		
XX			
SQ	Sequence 465 AA;		
<hr/>			
Query Match 66.8%; Score 1757.5; DB 20; Length 465;			

XX	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;	
KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;	
KW	schizophrenia; splice variant.	
XX		
OS	Cavia cobaya.	
XX		
PN	WO9947535-A1.	
XX		
XX		
PD	23-SEP-1999.	
XX		
PF	16-MAR-1999; 99WO-US05606.	
XX		
PR	16-MAR-1998; 98US-0078044.	
PR	05-JUN-1998; 98US-0088312.	
PR	28-OCT-1998; 98US-0106056.	
PR	22-JAN-1999; 99US-0236134.	
XX		
PA	(MILL-) MILLENNIUM PHARM INC.	
PA	(REGC) UNIV CALIFORNIA.	
PI	Chen H, Freimer NB;	
XX		
XX	WPI; 1999-562047/47.	
DR	N-PSDB; AA210756.	
DR		
XX		
PT	New HKNG1 polynucleotides useful in diagnosis and treatment of	
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and	
PT	schizophrenia -	
XX		
PS	Claim 1; Fig 8A-B; 205pp; English.	
XX		
XX	The present sequence is encoded by a guinea pig HKNG1 ortholog	
CC	gphkng1815 splice variant gphkng7c. HKNG1 (Hong Kong new gene 1)	
CC	is a gene associated with bipolar affective disorder (BAD). HKNG1	
CC	polynucleotides are useful to identify compounds modulating HKNG1	
CC	gene expression or HKNG1 polypeptide expression/activity. Compounds	
CC	inhibiting or enhancing HKNG1 gene expression or activity in	
CC	individuals can then be administered therapeutically to treat	
CC	HKNG1-mediated disorders, especially neuropsychiatric disorders	
CC	e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as	
CC	early-onset autosomal dominant myopia. The polynucleotides can be used	
CC	in gene therapy techniques to treat such disorders. They are also useful	
CC	in diagnosis to identify individuals having, or at risk of developing,	
CC	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such	
CC	mutations especially result in the production of a protein with a	
CC	different sequence to the human full-length HKNG1 polypeptide or	
CC	splice variant sequences, especially the substitution of a lysine for	
CC	a glutamic acid at residue 202 or 184. The polynucleotides are also	
CC	useful in gene mapping, to produce probes or primers to identify	
CC	similar sequences (e.g. mutants or sequences from different species)	
CC	and to produce transgenic animals.	
XX		
SQ	Sequence 450 AA;	

Query Match	59.2%;	Score 1558;	DB 20;	Length 450;
Best Local Similarity	63.9%;	Pred. No. 1.7e-120;		
Matches 299;	Conservative 62;	Mismatches 87;	Indels 20;	Gaps 3;

QY	30	MKPPPLVFIICLLWLKDSHCAPTWKDKTAISENLKSFSEVGEDADEEVKKAALTGTQMK	89
Db	1	mkpplmfvcvllwkdcapcwkdktaisenanfseagelvdgcvkialigkqm	60
QY	90	IMMERKEHTNLMSTLKKCREQBALKLLNEVQHEERLCSRESLADSGECRSL	149
Db	61	immerreeehsklmktlkkckeekgealklmnevhehleeeelslcqvsldswdecrac	120
QY	150	ENNCMIYTYTCQPSWSVKNKIERFRKIYQFLFPHEHNEKDLPISEKLIEDAQLTQM	209
Db	121	ensncmfddctcpawsvsknm-----endrsggpvskgvteadaqvshi	163
QY	210	EDVFSQLTVDVNSLNFNSFNVFMQOEQDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA	269

Db	164	ehvfnlsadvtlfnrslyvfkgirrefdqafqsgvftsgtadvcepfiffpslskepayra	223
QY	270	DLEQCWDIPNFQFCNFVSIIYESYSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ	329
Db	224	daepsaipnvnfqlcnlsfsvygsvekittlratdppkdkdsnggppiskilpeq	283
QY	330	DRGLCGELDONLSRCFKFHEKCKCQAHLSDECDPVPALHTEDEAIRLVNVSNOQYQGI	389
Db	284	drsgdkglnlsdcvnrfrckqcdqyisddcpnvypelyrelnealrlvsranqydvq	343
QY	390	LOMTRKHLEDTAYLVEKMRGQFCGVSELANQAPETELIFNSIOVVPRI--HEGNISQODE	447
Db	344	vmqtnhiedotllmekmreqfgvwselayqspgaedifnpvkvmlvaisahnegnsdqqd	403
QY	448	TMMTDLISLPSSNFTLKIPLEESAESSNFYGVVAKALQHFKEHFKTW 495	
Db	404	ttvps-sllpsnftlsspleksagnanfidhvvveklqhfkehfkwtw 450	
RESULT	8		
AA	AA30789		
ID	AA30789 standard; Protein; 374 AA.		
XX			
AC	AA30789;		
XX			
DT	23-NOV-1999 (first entry)		
XX			
DE	Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7c.		
XX			
KW	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;		
KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;		
KW	schizophrenia; splice variant.		
OS	Cavia cobaya.		
PN	WO9947535-A1.		
XX			
PD	23-SEP-1999.		
XX			
PF	16-MAR-1999; 99WO-US05606.		
XX			
PR	16-MAR-1998; 98US-0078044.		
PR	05-JUN-1998; 98US-0088312.		
PR	28-OCT-1998; 98US-0106056.		
PR	22-JAN-1999; 99US-0236134.		
XX			
PA	(MILL-) MILLENNIUM PHARM INC.		
PA	(REGC) UNIV CALIFORNIA.		
XX			
PI	Chen H, Freimer NB;		
XX			
DR	WPI; 1999-562047/47.		
DR	N-PSDB; AA210757.		
XX			
PT	New HKNG1 polynucleotides useful in diagnosis and treatment of		
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and		
PT	schizophrenia -		
XX			
PS	Claim 1; Fig 9A-B; 205pp; English.		
XX			
CC	The present sequence is encoded by a guinea pig HKNG1 ortholog		
CC	gphkng1815 splice variant gphkng7c. HKNG1 (Hong Kong new gene 1)		
CC	is a gene associated with bipolar affective disorder (BAD). HKNG1		
CC	polynucleotides are useful to identify compounds modulating HKNG1		
CC	gene expression or HKNG1 polypeptide expression/activity. Compounds		
CC	inhibiting or enhancing HKNG1 gene expression or activity in		
CC	individuals can then be administered therapeutically to treat		
CC	HKNG1-mediated disorders, especially neuropsychiatric disorders		
CC	e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as		
CC	early-onset autosomal dominant myopia. The polynucleotides can be used		
CC	in gene therapy techniques to treat such disorders. They are also useful		
CC	in diagnosis to identify individuals having, or at risk of developing,		
CC	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such		
CC	mutations especially result in the production of a protein with a		
CC	different sequence to the human full-length HKNG1 polypeptide or		
CC	splice variant sequences, especially the substitution of a lysine for		
CC	a glutamic acid at residue 202 or 184. The polynucleotides are also		
CC	useful in gene mapping, to produce probes or primers to identify		
CC	similar sequences (e.g. mutants or sequences from different species)		
CC	and to produce transgenic animals.		
XX			

Db 267 vqmtqyhledttllmekmreqfgwvselayqspgaedifnpkvkmvalsahegnssdqdd

QY	448	TMMTDLSPSSNTLKIDLESAESSNFEGYVAKALQHFKEHFKTW	495
DB	327	tvvps-sllpsnftlsspleksagnanfidhvvkvlqhfkefkfkw	373
RESULT 10			
AA	AY30794		
ID	AA30794	standard; Protein; 521 AA.	
AC	AA30794;		
DT	23-NOV-1999	(first entry)	
XX	Amino acid sequence of human HKNG1 splice variant HKNG1-delta7.		
XX	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;		
KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;		
KW	schizophrenia; splice variant.		
XX	Homo sapiens.		
OS			
PN	WO9947535-A1.		
XX			
PD	23-SEP-1999.		
XX	16-MAR-1999;	99WO-US05606.	
XX			
PR	16-MAR-1998;	98US-0078044.	
PR	05-JUN-1998;	98US-0088312.	
PR	28-OCT-1998;	98US-0106056.	
PR	22-JAN-1999;	99US-0236134.	
XX	(MILL-) MILLENNIUM PHARM INC.		
PA	(REGC) UNIV CALIFORNIA.		
XX			
PI	Chen H, Freimer NB;		
XX			
DR	WPI; 1999-562047/47.		
DR	N-PSDB; AA210762.		
XX			
PT	New HKNG1 polynucleotides useful in diagnosis and treatment of		
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and		
PT	schizophrenia		
XX			
PS	Claim 1; Fig 18A-B; 205pp; English.		
XX			
CC	The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)		
CC	splice variant. HKNG1 is a gene associated with bipolar affective		
CC	disorder (BAD). HKNG1 polynucleotides are useful to identify compounds		
CC	modulating HKNG1 gene expression or HKNG1 polypeptide expression/		
CC	activity. Compounds inhibiting or enhancing HKNG1 gene expression or		
CC	activity in individuals can then be administered therapeutically to		
CC	treat HKNG1-mediated disorders, especially neuropsychiatric disorders		
CC	e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as		
CC	early-onset autosomal dominant myopia. The polynucleotides can be used		
CC	in gene therapy techniques to treat such disorders. They are also useful		
CC	in diagnosis to identify individuals having, or at risk of developing,		
CC	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such		
CC	mutations especially result in the production of a protein with a		
CC	different sequence to the human full-length HKNG1 polypeptide or		
CC	splice variant sequences, especially the substitution of a lysine for		
CC	a glutamic acid at residue 202 or 184. The polynucleotides are also		
CC	useful in gene mapping, to produce probes or primers to identify		
CC	similar sequences (e.g. mutants or sequences from different species)		
CC	and to produce transgenic animals.		
XX			
SQ	Sequence 521 AA;		
Query Match 35.2%; Score 925.5; DB 20; Length 521;			
Best Local Similarity 77.0%; Pred. No. 4.8e-68;			
Matches 181; Conservative 7; Mismatches 16; Indels 31; Gaps 3;			

Qy	1	MKIAEKNEGSRSSWQLHWGDIANNNGNKKPPLLVIIVCLLWLKDSHCAPTWKDKTATS	60
Db	94	mkiaeknegpsrswqlhwgdiannsgnmkppllviivclllwkdshcaptwkdktais	153
Qy	61	ENLKSPSEVGEIDADAEVKKALTGIQKIMMERKEKEHTNLMSTLKKCREEKQAEALKLL	120
Db	154	enlkfsevgeidadeevkkaigtikqkimmerkekehtnltmstlkkcreekqealkll	213
Qy	121	NEVQEHLEEEERLCRESLADSWGECRSCLNENCMRIYTTTCQPSWSSVKNKI--ERFFERKI	178
Db	214	nevqehleeeerlcresladswwgecrsclenncmriytttcqpswssvknklltteeafqrc	273
Qy	179	YQFLPFPHEDNEKDLPISEKLIIEEDAQLTQMEDVFSQLT---VDVNSLFNRFSNV	230
Db	274	y-----lgrtedcvgnltricqgdvsnfmknaknv	302
RESULT 11			
AAU28048	ID	AAU28048 standard; Protein; 449 AA.	
XX	AC	AAU28048;	
XX	DT	18-DEC-2001 (first entry)	
XX	DE	Novel human secretory protein, Seq ID No 217.	
XX	KW	Human; secreted protein; arthritis; Crohn's disease; sepsis; shock; ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy; transgenic animal; Alzheimer's disease; Parkinson's disease; burn; amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia; ulcer; osteoporosis; bone degenerative disorder; periodontal disease; gut protection; lung; liver fibrosis; immune deficiency; infection; severe combined immunodeficiency; SCID; autoimmune disorder; allergy; multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma; fertility; analgesic; pain; antigen.	
XX	OS	Homo sapiens.	
XX	PN	WO200166689-A2.	
XX	PD	13-SEP-2001.	
XX	PF	05-MAR-2001; 2001WO-US04942.	
XX	PR	07-MAR-2000; 2000US-0519705.	
PR		19-MAY-2000; 2000US-0574454.	
PR		17-JUN-2000; 2000US-0596193.	
PR		14-JUL-2000; 2000US-0616847.	
PR		19-SEP-2000; 2000US-0665363.	
PR		20-OCT-2000; 2000US-0693267.	
XX	PA	(HYSE-) HYSEQ INC.	
XX	Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;		
PI	Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;		
XX	WPI; 2001-589934/66.		
DR	N-PSDB; AAS44948.		
XX	Novel polypeptides and nucleic acids obtained from cDNA libraries prepared from various human tissues, for diagnosis and treatment of cancer, neurological, inflammatory, and autoimmune disorders -		
XX	Example 3; SEQ ID No 217; 107pp; English.		
XX	The invention relates to novel isolated human secreted polypeptides (I) and polynucleotides (II). (I) and (II) are useful for treating inflammatory conditions such as arthritis, nephritis, Crohn's disease, ischaemia-reperfusion injury, shock, sepsis, immune responses, and is involved in increasing haematopoiesis, stem cell survival, bone growth and remodeling. (I), (II) and modulators of (II) are useful for prophylaxis or treatment of one or more cancers. (II) is also useful for		

CC creating transgenic animals useful for studying the in vivo activities of
 CC the polypeptide as well as for studying modulators of the polypeptides.
 CC (1) induces the proliferation of neural cells and regeneration of nerve
 CC and brain tissue and is useful for the treatment of central and
 CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,
 CC Parkinson's disease, Huntington's disease, and amyotrophic lateral
 CC sclerosis. In addition, (1) is involved in chemotactic or chemokinetic
 CC activity, regulation of haematopoiesis and is useful for treating myeloid
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
 CC tissue growth, and in tissue repair, healing of burns, incisions,
 CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative
 CC disorders, or periodontal disease. Furthermore, (1) is also useful for
 CC gut protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues, various immune deficiencies and
 CC disorders including severe combined immunodeficiency (SCID), bacterial or
 CC fungal infections, autoimmune disorders e.g. multiple sclerosis,
 CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic
 CC reactions and conditions, such as asthma or other respiratory problems.
 CC In addition, (1) affects biorhythms or circadian cycles of rhythms,
 CC fertility, metabolism, catabolism, anabolism, storage or elimination of
 CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides
 CC analgesic effects or other pain reducing effects, immunoglobulin like
 CC activity and can act as an antigen in a vaccine composition to raise an
 CC immune response. AAU28020-AAU28395 represent novel human secreted protein
 CC amino acid sequences of the invention.

Sequence 449 AA;

Query Match 16.5%; Score 434; DB 22; Length 449;
 Best Local Similarity 25.8%; Pred. No. 1.9e-27;
 Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

QY 30 MKPPLVFLVCLLKLKSHCAPTKW-----DKTAISENLKSFSEVGEIDAEVKKALT 83
 DB 1 mmktllivglll-----twesgvgldqtvsdnelqemsnqgskynkeiqnavn 51
 QY 84 GIKQMKIMMERKEKHTNLMSTLKKREKQKALKNLNEVQHEERLRESLADSWG 143
 DB 52 gvkqiktlietneerktilnleeeakkkedalnetresetkklpgvcnetmmalwe 111
 QY 144 ECRSCLNNCHRIYT-TCQPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLEE 202
 DB 112 eckpcikqtkcmkfyarvcrgsglvrgrleeflngsspfywmngdr-----idsilen 165
 QY 203 DAQLTQMEDV----FSQLTVDVNSLNFNSFNVFRQMOQEFDTQFQSHFIS-DTDLTEPYF 257
 DB 166 drqqtthldvmqdhfstrassildefqdrf-----ftrepqdy--hylpfslphrrphf 218
 QY 258 FPAFSKEPTWADLE-QCWDIPNFFQLFCNFSVSIYESVSITITKMLKAIEDLPKQDKAP 316
 DB 219 f--fpksxivrlmpfspyepnlhamqpflemihe-----qqamdihfspaf 267
 QY 317 DHGGLISKMLPGQDRGLGGLDQNLSCFKFEKCKOKCOAHLSEDC----PDVPALHTEL 372
 DB 268 qhpptefiredddrtvcrelhrnstgclrmkqdcckorellsvdcstnmpsqaklrrl 327
 QY 373 DEARLVNVSQQQGIQLQTRKLEDTAYLVERMRGQFGWVSELANOAPETEIFNSIQ 432
 DB 328 deslqvaerlkrkynellksyqwmntssilleqneqfnvwsrlantqgedqyirvt 387
 QY 433 VVPRIHEGNISKQDETMTDLSILPSSNFTLKIPLIESAESSNFIYGVAKALQHF-KEH 491
 DB 388 tvas-htsdsdpsvgvtevvvklfidsdpitvtvpvevskrpkfmetvaealqeyrkkh 446

RESULT 12

AA111704

ID AA111704 standard; Protein: 448 AA.

XX

AC AA111704;

XX

DT 20-JUN-1991 (first entry)
 XX Cytolysis inhibitor.
 DE Cytolysis inhibitor.
 XX Cytolysis inhibitor; perforin; immunological effector molecule;
 KW infertility.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT Protein /label= signal peptide
 FT Protein 22..226
 FT Protein /label= A-chain
 FT Protein 227..448
 FT Protein /label= B-chain
 XX DE3933850-A.
 PN 18-APR-1991.
 XX 06-OCT-1989; 89DE-3933850.
 XX 06-OCT-1989; 89DE-3933850.
 PR (SCHD) SCHERING AG.
 XX Tschopp J, Jenne D;
 XX WPT; 1991-118338/17.
 DR DNA sequence coding for cytolysis inhibitor - is strong inhibitor
 XX of terminal complement protein, eg perforin secreted by killer
 PT cells
 PS Claim 13; Page 9; 15pp; German.
 XX This cytolysis inhibitor is encoded by a 1.7kb BamHI-KpnI fragment
 CC isolated from a liver-specific cDNA library. It is a blood plasma
 CC component that inhibits immunological effector molecules. It is used
 CC for systemic or local treatment of inflammatory or autoimmune
 CC diseases mediated by complement or killer cells. It can also be
 CC used for detoxification of membrane-active and cytolytic proteins
 CC released by bacteria, fungi and insect venoms. The protein has a
 CC further use in the treatment of infertility caused by deficiency of
 CC cytolysis inhibitor. Monoclonal antibodies directed against the used
 CC cytolysis inhibitor are also covered by the invention. They are used
 CC to isolate or quantify the natural protein in human plasma.
 CC See also AAQ11501 and AAQ11502.

Sequence 448 AA;

Query Match 16.4%; Score 432; DB 12; Length 448;
 Best Local Similarity 25.8%; Pred. No. 2.7e-27;
 Matches 123; Conservative 96; Mismatches 205; Indels 52; Gaps 14;

QY 34 LLFVIVCLLWLKDSHCAPTWK-----DKTAISENLKSFSEVGEIDAEVKKALTGIKQ 87
 DB 4 lllfviglll-----twesgvgldqtvsdnelqemsnqgskynkeiqnavngvkq 54
 QY 88 MKIMMERKEKHTNLMSTLKKREKQKALKNLNEVQHEERLRESLADSWGECRS 147
 DB 55 iktlietneerktilnleeeakkkedalnetresetkklpgvcnetmmalweeckp 114
 QY 148 CLENNCHRIYT-TCQPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLEEADQL 206
 DB 115 clkqtkcmkfyarvcrgsglvrgrleeflngsspfywmngdr-----idsilendrqq 168
 QY 207 TOMEDV----FSQLTVDVNSLNFNSFNVFRQMOQEFDTQFQSHFIS-DTDLTEPYFPAP 261
 DB 169 thldvmqdhfstrassildefqdrf-----ftrepqdy--hylpfslphrrphf--f 219

Qy	147	SCLENNCMRIYT-TCQPSWSVKNKIRPRKRYQFLFPFHEDNEKDLPISEKLLIEDAQ	203
Db	115	pcLqtcmkfyarvcrgsglvgthgleefinqspfywingdr-----idslmendrql68	
Qy	206	LQMEDVFSQLTVDVNLSLFNRSNVFRMQOEFDQTFQSHFIS---DIDLTEPY-----256	
Db	169	qshvmdl-----medsfirasni-----mdelfqdrffnrepdfqtfspfgsshr214	
Qy	257	---PPFAFSKEPMTKADLEQCWDIPNPFQFLCNFVSIVESVSETITKMLKAIEDLPKOD313	
Db	215	gslfnpkrsfarnimpflftdl-nyhdmfqf---fdmihgaqamdaahlhripyh-268	
Qy	314	KAPHGGLISKMLP---GQDRGLGELDONLSRQFKHEKCKQCOAHLSEDC-----PDVPA367	
Db	269	-fpeag-----vpensndravckefrhnstgclrmkdqcekreillsvdcasnsqmq321	
Qy	368	LHTELDFAIRLVNVSNOOYGOILQMTKRKHLEDYALVEKMRGFGVWVSELANOAPETEI427	
Db	322	lrqelytslmaefsklydqglqsygqkmlntsslikqnegfswvswqlanltqnddry381	
Qy	428	FNSIQVVPRIHEGNISKQDETMTFDLSILPSSNFTLKIPLEESAESSNFTGYVAKALQH487	
Db	382	ylqvttv-nshgspdvspgltkvvklfdsyptitliipqevs--dpkfmetsvaealqq438	
Qy	488	FKE490	
Db	439	yrq441	
RESULT 15			
AAE03764	ID	AAE03764 standard; Protein; 363 AA.	
XX	AC	AAE03764;	
XX	AC	AAE03764;	
DT	07-AUG-2001	(first entry)	
DE	Human gene 1	encoded secreted protein HOF0C33, SEQ ID NO:34.	
KW	Human;	secreted protein; proliferative disorder; cancer; tumour;	
KW	fetal abnormality; development	abnormality; haematopoietic disorder;	
KW	immune system disorder; AIDS;	autoimmune disease; rheumatoid arthritis;	
KW	inflammation; allergy;	neurological disorder; Alzheimer's disease;	
KW	Parkinson's disease;	cognitive disorder; schizophrenia; asthma;	
KW	skin disorder; psoriasis;	sepsis; diabetes; atherosclerosis;	
KW	cardiovascular disorder;	angiogenic disorder; kidney disorder;	
KW	gastrointestinal disorder;	pregnancy-related disorder; gene therapy;	
KW	endocrine disorder; infection;	wound healing; vulnery;	
KW	cell culture; chemotaxis;	food additive;	
KW	binding partner	identification.	
OS	Homo sapiens.		
XX	Key	Location/Qualifiers	
FH	Peptide	1..21	
FT	Protein	/label= signal_peptide	
FT		22..363	
FT		/note= "Mature secreted protein"	
XX	WO200132837-Al.		
XX	10-MAY-2001.		
PD	17-OCT-2000;	2000WO-US28664.	
PF	02-NOV-1999;	99US-0163085.	
PR	17-DEC-1999;	99US-0172411.	
XX	(HUMA-) HUMAN GENOME	SCI INC.	
PA	Fiscella M, Wei P, Lafleur DW, Olsen HS, Baker K, Ebner R;		
PI	Komatsoulis G, Rosen CA, Ruben SM, Duan RD, Young PE, Florence K		
PI	Moore PA, Birse CE, Ni J, Soppet DR, Shi Y;		

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:33:26 ; Search time 24 Seconds
(without alignments)
503.778 Million cell updates/sec

Title: US-09-722-544A-2
Perfect score: 2632
Sequence: 1 MKIKAEKNEGSRWQLHW.....FIGYVAKALQHFKEHFTKW 495

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2632	100.0	495	4	US-09-268-992-2
2	2464	93.6	466	4	US-09-268-992-64
3	2350	89.3	446	4	US-09-268-992-51
4	2335	88.7	477	4	US-09-268-992-4
5	1755.5	66.7	465	4	US-09-268-992-49
6	1639	62.3	466	4	US-09-268-992-39
7	1558	59.2	450	4	US-09-268-992-41
8	1277	48.5	374	4	US-09-268-992-43
9	1272.5	48.3	373	4	US-09-268-992-45
10	921	35.0	208	4	US-09-268-992-67
11	145	5.5	1388	2	US-08-685-576-1
12	141	5.4	1388	2	US-08-685-576-4
13	130.5	5.0	2482	1	US-08-328-254-6
14	129	4.9	816	2	US-08-533-306A-6
15	129	4.9	816	2	US-08-742-923A-6
16	128.5	4.9	3248	5	US-08-353-700-1
17	128.5	4.9	3248	5	PCT-US95-16216-1
18	128	4.9	885	2	US-08-533-306A-4
19	128	4.9	885	2	US-08-742-923A-4
20	125.5	4.8	1354	3	US-08-685-871-2
21	121	4.6	828	2	US-08-993-228-21
22	121	4.6	1618	1	US-07-853-913-4
23	120	4.6	2154	2	US-08-841-349-4
24	119.5	4.5	435	2	US-08-531-439B-4
25	119.5	4.5	3111	2	US-08-460-309-4
26	119.5	4.5	3111	2	US-08-125-077-4
27	118	4.5	1886	4	US-08-938-105-3

ALIGNMENTS

RESULT 1

US-09-268-992-2
; Sequence 2, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-2

Query Match 100.0%; Score 2632; DB 4; Length 495;
Best Local Similarity 100.0%; Pred. No. 1.9e-244; Indels 0; Gaps 0;
Matches 495; Conservative 0; Mismatches 0;
Qy 1 MKIKAEKNEGSRWQLHWGDIANNNGMKPPLLVFVCLLWLDKSHCAPTWKDKTAIS 60
Db 1 MKIKAEKNEGSRWQLHWGDIANNNGMKPPLLVFVCLLWLDKSHCAPTWKDKTAIS 60
Qy 61 ENLKSFEVGEIDAEVKKALGTGKIMMERKEHTNLMSTLKKCREKQEAALKL 120
Db 61 ENLKSFEVGEIDAEVKKALGTGKIMMERKEHTNLMSTLKKCREKQEAALKL 120
Qy 121 NEVOEHLEERLCRESLADSWGECRSCLENNCMRIYTCOPSSVKNKTERFRKIYQ 180
Db 121 NEVOEHLEERLCRESLADSWGECRSCLENNCMRIYTCOPSSVKNKTERFRKIYQ 180
Qy 181 FLFFPHEDNEKDLPISEKLIEDAQLTQMEDVFSOLTVDVNSLFRNSRNFVRQMQEFDQ 240
Db 181 FLFFPHEDNEKDLPISEKLIEDAQLTQMEDVFSOLTVDVNSLFRNSRNFVRQMQEFDQ 240
Qy 241 TFQSHFISDTDLTEPYFPFPAFSKEPMTKADLEQCWDIPNLFQFCNFSVSIYESVSETIT 300

Db 241 TFSHFISDITLTPYFFAFKSKPTWKADLEQCNDIPNFQFQCFNFSVIVSVSETIT 300
QY 301 KMLKAIEDLPKQKADPHGGLSKMLPGQDRGLCGELDNLSRCFKFHEKCKQCAHLSE 360
Db 301 KMLKAIEDLPKQKADPHGGLSKMLPGQDRGLCGELDNLSRCFKFHEKCKQCAHLSE 360
QY 361 DCPDVPALHTLDEAIRLVNSNOQYGOILOMTRKHLEDATAYLVKMRGQFGWVSELANQ 420
Db 361 DCPDVPALHTLDEAIRLVNSNOQYGOILOMTRKHLEDATAYLVKMRGQFGWVSELANQ 420
QY 421 APETEIFNSIOVVPRIHEGNTSKQDETMTDLSILPSSNFTLKIPLESASSNFYGV 480
Db 421 APETEIFNSIOVVPRIHEGNTSKQDETMTDLSILPSSNFTLKIPLESASSNFYGV 480
QY 481 VAKALQHFKHEFKTW 495
Db 481 VAKALQHFKHEFKTW 495

RESULT 2

US-09-268-992-64

; Sequence 64, Application US/09268992

; Patent No. 6342351

; GENERAL INFORMATION:

; APPLICANT: Chen, H.

; APPLICANT: Freimer, N.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING

; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS

; FILE REFERENCE: 7853-138

; CURRENT APPLICATION NUMBER: US/09/268,992

; CURRENT FILING DATE: 1999-03-16

; EARLIER APPLICATION NUMBER: 09/236,134

; EARLIER FILING DATE: 1999-01-22

; EARLIER APPLICATION NUMBER: 60/106,056

; EARLIER FILING DATE: 1998-10-28

; EARLIER APPLICATION NUMBER: 60/088,312

; EARLIER FILING DATE: 1998-06-05

; EARLIER APPLICATION NUMBER: 60/078,044

; EARLIER FILING DATE: 1998-03-16

; NUMBER OF SEQ ID NOS: 84

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 64

; LENGTH: 466

; TYPE: PRT

; ORGANISM: Homo sapiens

; ORGANISM: Homo sapiens

US-09-268-992-64

Query Match 93.6%; Score 2464; DB 4; Length 466;
Best Local Similarity 100.0%; Pred. No. 2.4e-228; Mismatches 0; Indels 0; Gaps 0;
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 MKPPLLVFVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEDADAEVVKKALTGKQMK 89
Db 1 MKPPLLVFVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEDADAEVVKKALTGKQMK 60
QY 90 IMMERKEHTNLMSTYKCKREKQKALKNVEQVHLEERLCRESLADSWGECRSL 149
Db 61 IMMERKEHTNLMSTYKCKREKQKALKNVEQVHLEERLCRESLADSWGECRSL 120
QY 150 ENNCRIYTTCPQSWSSVSNKRIERFRKIYQFLFPFHEDNEKDLPISEKLIIEADAQLTQM 209
Db 121 ENNCRIYTTCPQSWSSVSNKRIERFRKIYQFLFPFHEDNEKDLPISEKLIIEADAQLTQM 180
QY 210 EDVFSOLTVDVNSLNRSENFVRMQQEDQTFQSHFISDITLTPYFFAFKSKPTWK 269
Db 181 EDVFSOLTVDVNSLNRSENFVRMQQEDQTFQSHFISDITLTPYFFAFKSKPTWK 240
QY 270 DLEQCWDIPNFQFCNFSVSVSETITKMLKAIEDLPKQKADPHGGLSKMLPGQ 329
Db 241 DLEQCWDIPNFQFCNFSVSVSETITKMLKAIEDLPKQKADPHGGLSKMLPGQ 300

QY 330 DRGLCGELDNLSRCFKFHEKCKQCAHLSEDCPDVPALHTLDEAIRLVNSNOQYGOI 389
Db 301 DRGLCGELDNLSRCFKFHEKCKQCAHLSEDCPDVPALHTLDEAIRLVNSNOQYGOI 360
QY 390 LOWTRKHLEDATAYLVKMRGQFGWVSELANQAPETEIFNSIQVVPRIHEGNTSKODETM 449
Db 361 LOWTRKHLEDATAYLVKMRGQFGWVSELANQAPETEIFNSIQVVPRIHEGNTSKODETM 420
QY 450 MTDSLILPSSNFTLKIPLESASSNFYGVYVAKALQHFKHEFKTW 495
Db 421 MTDSLILPSSNFTLKIPLESASSNFYGVYVAKALQHFKHEFKTW 466

RESULT 3

US-09-268-992-51

; Sequence 51, Application US/09268992

; Patent No. 6342351

; GENERAL INFORMATION:

; APPLICANT: Chen, H.

; APPLICANT: Freimer, N.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING

; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS

; FILE REFERENCE: 7853-138

; CURRENT APPLICATION NUMBER: US/09/268,992

; CURRENT FILING DATE: 1999-03-16

; EARLIER APPLICATION NUMBER: 09/236,134

; EARLIER FILING DATE: 1999-01-22

; EARLIER APPLICATION NUMBER: 60/106,056

; EARLIER FILING DATE: 1998-10-28

; EARLIER APPLICATION NUMBER: 60/088,312

; EARLIER FILING DATE: 1998-06-05

; EARLIER APPLICATION NUMBER: 60/078,044

; EARLIER FILING DATE: 1998-03-16

; NUMBER OF SEQ ID NOS: 84

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 51

; LENGTH: 446

; TYPE: PRT

; ORGANISM: Homo sapiens

; ORGANISM: Homo sapiens

US-09-268-992-51

Query Match 89.3%; Score 2350; DB 4; Length 446;
Best Local Similarity 100.0%; Pred. No. 2e-217; Mismatches 0; Indels 0; Gaps 0;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 APTWKDKTAISENLKSFSEVGEDADAEVVKKALTGKQMKIMMERKEHTNLMSTLKKC 109
Db 1 APTWKDKTAISENLKSFSEVGEDADAEVVKKALTGKQMKIMMERKEHTNLMSTLKKC 60
QY 110 REEKQKALKNVEQVHLEERLCRESLADSWGECRSCLENNCMRIYTTCPQSWSSVKN 169
Db 61 REEKQKALKNVEQVHLEERLCRESLADSWGECRSCLENNCMRIYTTCPQSWSSVKN 120
QY 170 KIERFRKIYQFLFPFHEDNEKDLPISEKLIIEADAQLTQMEDVFSOLTVDVNSLNRSEN 229
Db 121 KIERFRKIYQFLFPFHEDNEKDLPISEKLIIEADAQLTQMEDVFSOLTVDVNSLNRSEN 180
QY 230 VFRMQQEDQTFQSHFISDITLTPYFFAFKSKPTWKADLEQCWDIPNFQFCNFSV 289
Db 181 VFRMQQEDQTFQSHFISDITLTPYFFAFKSKPTWKADLEQCWDIPNFQFCNFSV 240
QY 290 STYESVSETITKMLKAIEDLPKQKADPHGGLSKMLPGQDRGLCGELDNLSRCFKPHE 349
Db 241 STYESVSETITKMLKAIEDLPKQKADPHGGLSKMLPGQDRGLCGELDNLSRCFKPHE 300
QY 350 KCKQCAHLSEDCPDVPALHTLDEAIRLVNSNOQYGOILOMTRKHLEDATAYLVKMRG 409
Db 301 KCKQCAHLSEDCPDVPALHTLDEAIRLVNSNOQYGOILOMTRKHLEDATAYLVKMRG 360
QY 410 QFGWVSELANQAPETEIFNSIQVVPRIHEGNTSKODETMTDLSILPSSNFTLKIPLEE 469
Db 361 QFGWVSELANQAPETEIFNSIQVVPRIHEGNTSKODETMTDLSILPSSNFTLKIPLEE 420

Qy 470 SAESSNFIGYVAKALQHFKEHFKTW 495
|||||
Db 421 SAESSNFIGYVAKALQHFKEHFKTW 446

RESULT 4
US-09-268-992-4
; Sequence 4, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-4

Query Match 88.7%; Score 2335; DB 4; Length 477;
Best Local Similarity 91.2%; Pred. No. 6.3e-216;
Matches 448; Conservative 2; Mismatches 1; Indels 40; Gaps 2;
Qy 25 NNSGNMKPPLVIFVCLLWKDSHCAPTWKDKTAISENLKSFSEVGEIDADEEVKALTG 84
:|||||
Db 7 SMSGNKPPLVIFVCLLWKDSHCAPTWKDKSAISENLKSFSEVGEIDADEEVKALTG 66
:|||||
Qy 85 IKQKIMMERKEK-----EHTNLMSTLKKCKREEKQKALKLLNEVQ 124
|||||
Db 67 IKQKIMMERKEKANAPETEIFNSIQVVPRIEHTNLMSTLKKCKREEKQKALKLLNEVQ 126
|||||
Qy 125 EHLEEBERLCRESLADSGECRSCLENNCMRIYTTCCPSWSSVKNKIERFFRKIIYQFLFP 184
|||||
Db 127 EHLEEBERLCRESLADSGECRSCLENNCMRIYTTCCPSWSSVKNKIERFFRKIIYQFLFP 186
|||||
Qy 185 FHEDNEKDLPISEKLTIEEDAQLTQMEDVFSOLTVDVNSLNFNRSFNVFQMQQERDOTFS 244
|||||
Db 187 FHEDNEKDLPISEKLTIEEDAQLTQMEDVFSOLTVDVNSLNFNRSFNVFQMQQERDOTFS 246
|||||
Qy 245 HFISDTDLTEPYFFPAFSKEPMTKADLEQCDWIPNFOLFNCNFSYIYESVSETITKMLK 304
|||||
Db 247 HFISDTDLTEPYFFPAFSKEPMTKADLEQCDWIPNFOLFNCNFSYIYESVSETITKMLK 306
|||||
Qy 305 ATEDLPKQKAPDHPGGLISKMLPGDRLGCGELDONLSRCFKFHEKCKQKQAHLSDECPD 364
|||||
Db 307 ATEDLPKQKAPDHPGGLISKMLPGDRLGCGELDONLSRCFKFHEKCKQKQAHLSDECPD 366
|||||
Qy 365 VPALHTELDEARLVNSNQYGOQLQMTKRKLEDTAYLVEKMRGQFGWVSELANQAPET 424
|||||
Db 367 VPALHTELDEARLVNSNQYGOQLQMTKRKLEDTAYLVEKMRGQFGWVSEL----- 419
|||||
Qy 425 EIIFNSIQVVPRIHEGNISKQDETMMTDLISLPSSNFTLKIPLEESAESSNFIGYVAKA 484
|||||
Db 420 -----HEGNISKQDETMMTDLISLPSSNFTLKIPLEESAESSNFIGYVAKA 466
|||||
Qy 485 LQHFKEHFKTW 495
|||||

Db 467 LQHFKEHFKTW 477

RESULT 5
US-09-268-992-49
; Sequence 49, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 49
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Bos sp.
US-09-268-992-49

Query Match 66.7%; Score 1755.5; DB 4; Length 465;
Best Local Similarity 70.8%; Pred. No. 2.8e-160;
Matches 330; Conservative 58; Mismatches 77; Indels 1; Gaps 1;
Qy 30 MKPPLLVIFVCLLWKDSHCAPTWKDKTAISENLKSFSEVGEIDADEEVKALTGKQMK 89
|||||
Db 1 MKPPLLVIFVCLLWKDSHCAPTWKDKTSIREDPKGFSGAGEIDVDEEVKALIGKQMK 60
|||||
Qy 90 IMMERKEHEHNLMSTLKKCKREEKQKALKLLNEVQEHLEEBERLCRESLADSGECRSC 149
:|||||
Db 61 ILMERREHEHKLMTLKKCKREEKQKALKLLNEVQEHLEEBERLCRESLADSGECRSC 120
|||||
Qy 150 ENNCRIYTTCCPSWSSVKNKIERFFRKIIYQFLFPFHEDNEKDLPISEKLTIEEDAQLTQ 209
:|||||
Db 121 ESDCMRFTTCCQSSWSSMKSTIERVFRKIYQFLFPFHEDDEKELPVGKFTTEEDVQLMQ 180
|||||
Qy 210 EDVFSQLTVDVNSLNFNRSFNVFQMQQERDOTFSHFISDTDLTEPYFFPAFSKEPMTKA 269
:|||||
Db 181 ENVFSQLTVDVNSLNFNRSFNVFQMQQERDOTFSHFISDTDLTEPYFFPAFSKEPMTKA 240
|||||
Qy 270 DLEQCDWIPNFOLFNCNFSYIYESVSETITKMLKAIEDLPKQKAPDHPGGLISKMLPGQ 329
:|||||
Db 241 HPMQSWDIPSFQFLFCNFSYIYESVSETITKMLKAIEDLPKQKAPDHPGGLISKMLPGQ 300
|||||
Qy 330 DRGLGCGELDONLSRCFKFHEKCKQKQAHLSDECPDVPALHTELDEARLVNSNQYGO 389
|||||
Db 301 GRGLGCGEQNSSECLQFHARCQKQKQDYLWADCPAVPELYTKADEALELVNLSNQYQ 360
|||||
Qy 390 LQMTKRKLEDTAYLVEKMRGQFGWVSELANQAPETEIFNSIQVVPRIHEGNISKQDETMM 449
:|||||
Db 361 LQMTQHHLEDITYLMEKMRGQFGWVSELANQAPETEIFNSIQVVPRIHEGNISKQDETMM 420
|||||
Qy 450 MTDLSILPSSNFTLKIPLEESAESSNFIGYVAKALQHFKEHFKTW 495
:|||||
Db 421 I-DTSLPSSNFTLKIPLEESAESSNFIGYVAKALQHFKEHFKTW 465
|||||

RESULT 6
US-09-268-992-39
; Sequence 39, Application US/09268992.
; Patent No. 6342351

```

; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Cavia sp.
; US-09-268-992-39

Query Match      62.3%; Score 1639; DB 4; Length 466;
Best Local Similarity 66.7%; Pred. No. 4.4e-149;
Matches 312; Conservative 63; Mismatches 89; Indels 4; Gaps 3;

QY 30 MKPPLLVFIVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDADBEVKKALTGIKQM 89
Db 1 MKPLLPFPVCLLWLDKCHCAPTWKDKTAISENANSEAGEIDVDGEVKIALIGIKQM 60
QY 90 IMMERKEKHTNLMSTLKKCKEKEQKALKLLNEVQHELEERLCRESLADSGECRSL 149
Db 61 IMMERREEHSLKMLTKLKKCKEKEQKALKLMNEVHEHLEESLCOVSLADSWDECRACL 120
QY 150 ENNCMRIYTTTCQPSWSSVKNKIERFRKIIYQFLFPPHEDNEKDLPISEKLIBEDAQTOM 209
Db 121 ESNCRMRDFTTCQPAWSSVKNM-----ENDRSGPVSKGVTEEDAQVSHI 163
QY 210 EDVFSQTLTVDNLSFNRSFNVFRMQOEFDQTFQSHFISDITLDEPYFFPAFSKEPMTKA 269
Db 180 EHVSQLSADVTSFNRSFLVFKQLREFDQAFQSVFTSGTDTTEPFPFSLSKEPAYRA 239
QY 270 DLQOCWDIPNFFOLFCNFSVIYESVSETTKMKLKAIEDLPKODKAPDHGGLISKMLPGQ 329
Db 240 DABPSWAIPNVFOLLNLSFSVQSVSEKLTITLRATEDPPKODKDSNQGPGISKILPEQ 299
QY 330 DRGLGELDONLSRCFKFEKCKCOAHLSEDCPDVPAHTEDEAIRLVNVSNOQYQGI 389
Db 300 DRSGDKLGNLSDCVNFRKRCQKQDYLSDDCPNVPPELYRELNEALRLVRSRNOQYDQV 359
QY 390 LQWTRKHLEDYALVEKMRQFGWSELANQAPETEIFNSIQVWPRI--HEGNISKQDE 447
Db 360 VQMTQVHLEDYTLLEKMRQFGWSELAYQSPGAEDIFNPVKVMVALSAHEGNSDDDD 419
QY 448 TMTDLSILPSSNFTLKIPLEESAESSNFTGYVYVAKALQHFKEHFKTW 495
Db 420 TVVPS-SLLPSSNFTLSSPLEKSAGNANFDHVVVKVQLQHFKEHFKTW 466

RESULT 7
US-09-268-992-41
; Sequence 41, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Cavia sp.
; US-09-268-992-39
```

```

; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Cavia sp.
; US-09-268-992-41

Query Match      59.2%; Score 1558; DB 4; Length 450;
Best Local Similarity 63.9%; Pred. No. 2.6e-141;
Matches 299; Conservative 62; Mismatches 87; Indels 20; Gaps 3;

QY 30 MKPPLLVFIVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDADBEVKKALTGIKQM 89
Db 1 MKPLLPFPVCLLWLDKCHCAPTWKDKTAISENANSEAGEIDVDGEVKIALIGIKQM 60
QY 90 IMMERKEKHTNLMSTLKKCKEKEQKALKLLNEVQHELEERLCRESLADSGECRSL 149
Db 61 IMMERREEHSLKMLTKLKKCKEKEQKALKLMNEVHEHLEESLCOVSLADSWDECRACL 120
QY 150 ENNCMRIYTTTCQPSWSSVKNKIERFRKIIYQFLFPPHEDNEKDLPISEKLIBEDAQTOM 209
Db 121 ESNCRMRDFTTCQPAWSSVKNM-----ENDRSGPVSKGVTEEDAQVSHI 163
QY 210 EDVFSQTLTVDNLSFNRSFNVFRMQOEFDQTFQSHFISDITLDEPYFFPAFSKEPMTKA 269
Db 164 EHVSQLSADVTSFNRSFLVFKQLREFDQAFQSVFTSGTDTTEPFPFSLSKEPAYRA 223
QY 270 DLQOCWDIPNFFOLFCNFSVIYESVSETTKMKLKAIEDLPKODKAPDHGGLISKMLPGQ 329
Db 224 DABPSWAIPNVFOLLNLSFSVQSVSEKLTITLRATEDPPKODKDSNQGPGISKILPEQ 283
QY 330 DRGLGELDONLSRCFKFEKCKCOAHLSEDCPDVPAHTEDEAIRLVNVSNOQYQGI 389
Db 284 DRSGDKLGNLSDCVNFRKRCQKQDYLSDDCPNVPPELYRELNEALRLVRSRNOQYDQV 343
QY 390 LQWTRKHLEDYALVEKMRQFGWSELANQAPETEIFNSIQVWPRI--HEGNISKQDE 447
Db 344 VQMTQVHLEDYTLLEKMRQFGWSELAYQSPGAEDIFNPVKVMVALSAHEGNSDDDD 403
QY 448 TMTDLSILPSSNFTLKIPLEESAESSNFTGYVYVAKALQHFKEHFKTW 495
Db 404 TVVPS-SLLPSSNFTLSSPLEKSAGNANFDHVVVKVQLQHFKEHFKTW 450

RESULT 8
US-09-268-992-43
; Sequence 43, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
```

; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 43
; LENGTH: 374
; TYPE: PRG
; ORGANISM: Cavia sp.
US-09-268-992-43

Query Match 48.58; Score 1277; DB 4; Length 374;
Best Local Similarity 54.58; Pred. No. 1.9e-114;
Matches 255; Conservative 44; Mismatches 73; Indels 96; Gaps 3;

QY 30 MKPPLVFIIVCLLWLDKSHCAPTKWKDTAISENLKSFSEVGEIDADAEVKKALGKQMK 89
DB 1 MKPLLMFPVCLLWLDKCHCAPTKWKDTAISENANSEAGEIDVDGEVKIALGKQMK 60

QY 90 IMMERKEHTNLMSTLKKCKEKEQKALKLLNEVQEHLEEBERLCRESLADSWGECRSL 149
DB 61 IMMERREBEHSLKMTLKKCKEKEQKALKLLNEVQEHLEEBERLCRESLADSWGECRSL 120

QY 150 ENNCRIYTTCPSSWSSVKNKIERFRKYYQFLFPFHEDNEKDLPISEKLEEDAQLTQM 209
DB 121 ENSCMREDTTCOPAWSSVKN-----140

QY 210 EDVFSQLTVDVNSLNFNSFNVFQMQQEFDTQFSHFISDTDLTEPYFFPAFSKEPMTKA 269
DB 141 -----MPAYRA 147

QY 270 DLEOCWDIPNFQFPCNFVSIVSVSETITKMLKAIEDLPKQKADPHGLSLKMLPGQ 329
DB 148 DAEPWAIPNVFQLLNLCSFSVQSVSEKLTTLRATEDPPKQKDSNOGGPISKILPEQ 207

QY 330 DRGLCGELDONLRCFKFHEKCKQKQAHLSDCPDVPAHTELDIAIRLVNVSNOQYQOI 389
DB 208 DRGSDGKLQNLSDCVNFRKRCQKQDYLSDDCPNVPELYRELNEALRLVSRNQYQDV 267

QY 390 LQMTKRKLEDTAYLVERKMRGQGWVSELANQAPETEIFNSIQVVPRI--HEGNISKQDE 447
DB 268 VQMTQYHLEDITLLMEKMRQEGWVSELAYQSPCAEDIFNPVKVMVALSAHEGNSDQDD 327

QY 448 TMWTDLSLPSNTTLKPILESASSNFYGVYVAKALQHFKEHFKTW 495
DB 328 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVVKVQLQHFKEHFKTW 374

RESULT 9
US-09-268-992-45
; Sequence 45, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45
; LENGTH: 373
; TYPE: PRG

; ORGANISM: Cavia sp.
US-09-268-992-45

Query Match 48.38; Score 1272.5; DB 4; Length 373;
Best Local Similarity 54.58; Pred. No. 5.1e-114;
Matches 255; Conservative 45; Mismatches 71; Indels 97; Gaps 4;

QY 30 MKPPLVFIIVCLLWLDKSHCAPTKWKDTAISENLKSFSEVGEIDADAEVKKALGKQMK 89
DB 1 MKPLLMFPVCLLWLDKCHCAPTKWKDTAISENANSEAGEIDVDGEVKIALGKQMK 60

QY 90 IMMERKEHTNLMSTLKKCKEKEQKALKLLNEVQEHLEEBERLCRESLADSWGECRSL 149
DB 61 IMMERREBEHSLKMTLKKCKEKEQKALKLLNEVQEHLEEBERLCRESLADSWGECRSL 120

QY 150 ENNCRIYTTCPSSWSSVKNKIERFRKYYQFLFPFHEDNEKDLPISEKLEEDAQLTQM 209
DB 121 ENSCMREDTTCOPAWSSVKN-----140

QY 210 EDVFSQLTVDVNSLNFNSFNVFQMQQEFDTQFSHFISDTDLTEPYFFPAFSKEPMTKA 269
DB 141 -----MPAYRA 146

QY 270 DLEOCWDIPNFQFPCNFVSIVSVSETITKMLKAIEDLPKQKADPHGLSLKMLPGQ 329
DB 147 DAEPWAIPNVFQLLNLCSFSVQSVSEKLTTLRATEDPPKQKDSNOGGPISKILPEQ 206

QY 330 DRGLCGELDONLRCFKFHEKCKQKQAHLSDCPDVPAHTELDIAIRLVNVSNOQYQOI 389
DB 207 DRGSDGKLQNLSDCVNFRKRCQKQDYLSDDCPNVPELYRELNEALRLVSRNQYQDV 266

QY 390 LQMTKRKLEDTAYLVERKMRGQGWVSELANQAPETEIFNSIQVVPRI--HEGNISKQDE 447
DB 267 VQMTQYHLEDITLLMEKMRQEGWVSELAYQSPCAEDIFNPVKVMVALSAHEGNSDQDD 326

QY 448 TMWTDLSLPSNTTLKPILESASSNFYGVYVAKALQHFKEHFKTW 495
DB 327 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVVKVQLQHFKEHFKTW 373

RESULT 10
US-09-268-992-67
; Sequence 67, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 208
; TYPE: PRG
; ORGANISM: Homo sapiens
US-09-268-992-67

Query Match 35.08; Score 921; DB 4; Length 208;
Best Local Similarity 99.48; Pred. No. 1.2e-80;
Matches 170; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTKAEKNEGSPRSWOLHWGDIANNNGMKPPLLVFIVCLLWKDQSHCAPTKWCKTAIS 60
DB 33 MKTKAEKNEGSPRSWOLHWGDIANNNGMKPPLLVFIVCLLWKDQSHCAPTKWCKTAIS 92
QY 61 ENLKSFSEVEIDADEVKALGTGKQKIMMERKEKEHTNLMSTLKKCKREEKQKALKLL 120
DB 93 ENLKSFSEVEIDADEVKALGTGKQKIMMERKEKEHTNLMSTLKKCKREEKQKALKLL 152
QY 121 NEVQEHLEEEERLCRESLADSWGECRCLNENCMRIYTTCCPSWSSVKNKI 171
DB 153 NEVQEHLEEEERLCRESLADSWGECRCLNENCMRIYTTCCPSWSSVKNKL 203

RESULT 11
US-08-685-576-1
; Sequence 1, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Kozo
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-685-576-1

Query Match 5.5%; Score 145; DB 2; Length 1388;
Best Local Similarity 16.5%; Pred No. 8.1e-05;
Matches 94; Conservative 97; Mismatches 160; Indels 212; Gaps 20;

QY 54 KDKTAISENLKSFSEVEIDADEVKALGTGKQKIMMERKEKEHTNLMSTLKKCKREEKQKALKLL 104

DB 780 KQDVLNEDVRNLT-----LKIEQETQKRCLTQNDLAWQTOVNTLAKMSEKQLKQENHLL 835
QY 105 TLK-----KCREBKQKALKLLNEVQEHLEEE-----RLCR 135
DB 836 EMKMSLEKQNAELRKERQADGQMKELQDLQLEAEQYFSTLYKTQVRELKECEKTKLCK 895
QY 136 E-----SLADSWGECRCLNENCMRIYTTCCPSWSSVKNKIETFRFRKIYQFLFPFHED 189
DB 896 ELQOKKQELQDE-----RDSLAAQLEITLTAKDSEQLARSAEQQYSLEKE 942
QY 190 E--KDLPISE-----KLIEEDAQLTOMEDVFSQLTVDVNSLENRSFNVFQMOQERDQ 240
DB 943 KIMKELEIKEMMARHKQELTEKDATASLEETNRTLTSDVANLANEKEELNKKLKEAQO 1002
QY 241 TQSHFISDTDLTEPYFFPAFSKPEMTKADLEOCNDIPNFQFCNFSVIYESVSETIT 300
DB 1003 LSR---LKDEISAAAIKAQFEKQLLTERTLK-----TQAVN 1036
QY 301 KMLKATE-----DLPKQDKAPDHGGLISKMLPGQDRGLCGEL---DQNLSCF 345
DB 1037 KLAEMNRKPEVVRGNDTQVRRKEK-----ENRKLHMLKSEKREKLTQOMI 1082
QY 346 KPEHKQKQKQAHLSQDPVAPALHTDELDEAIRLVNVSQYQGIQLQMTKRKHL----- 398
DB 1083 KYQKELNEMQAQIAEE---SQIRIELOMTLDSKDSIDIEQLRSQALHIGLDSSSIGSG 1138
QY 399 -----DTAYLYVEKMRG-----QFGWVSEL----- 417
DB 1139 PGDTEADGCPESRLEGWLSLVRNNTKFGWVKYIVVSSKKILFYDSEQKQESNPYM 1198
QY 418 -----ANQAPETETIFNSIQVVPRI-----HEGNISKQDETMTMTDLSILPSSNFT 462
DB 1199 VLDIDKLPHVRPVTQDVYRADAKEIPRFQILYANEGESKKEQE----- 1243
QY 463 LKIPLEESAESNFI---GYVVAKALQHF 488
DB 1244 --FPVEPVGEKSNYICHKGHEFIPTLYHF 1270

RESULT 12
US-08-685-576-4
; Sequence 4, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Kozo
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-685-576-1

;; FILING DATE: 05-JAN-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 8-131206
;; FILING DATE: 26-APR-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Bent, Stephen A.
;; REGISTRATION NUMBER: 29,768
;; REFERENCE/DOCKET NUMBER: 16887/843
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)672-5300
;; TELEFAX: (202)672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1388 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-685-576-4

Query Match 5.4%; Score 141; DB 2; Length 1388;
Best Local Similarity 16.8%; Pred. No. 0.0002;
Matches 94; Conservative 97; Mismatches 176; Indels 192; Gaps 20;
QY 54 KDKTAISENLSFSEVGIDEAEVKKALGTGKMKI-----MMERKEKEHTNLS 104
DB 780 KOKOVNEDVRNLT-----LKTEQFQKRLTQNDLKMTQOVNLMKSEKQKQENHLM 835
QY 105 TLK-----KCREKEQALKLNEVQHLLEEE-----RLCRESIADSWGEC----- 145
DB 836 ENKMLEKQNAELRXERQADQOMKELQDLAEQYFSTLYKTQVRELKECEKTKLKG 895
QY 146 ---RSCLENNCHRYTTCQPSWSSVKNKIERFFRYQFLPFPHEDNE--KDLPISE--- 197
DB 896 ELQQRKQELQDRDSLAQALETITLKADSEQLARSIAEQYSDLEKEKIMKELEIKMA 955
QY 198 ----KLIEDAOLQMEDVFSOLTVDVNSLFNRS---FNVFQMOQEFDTQFSHFISDT 250
DB 956 RHQBELTEKDATIASLEETNRLTSDVANLANEKEELNKLKDVQEQLSR-----LKDE 1009
QY 251 DLTEPYFFPAPFKEPMTKADLQCCWDIPNFOLFNCNFSVYSVESVETITMKLKAIE--- 307
DB 1010 EISAAAIKAQFEKQLITERTLK-----TQAVNKLAEITMNRKE 1046
QY 308 -----DLPKQDAPDHGGLSKMLPCQDRGLGCEL-----DONLSRCFKFHEKCKCO 355
DB 1047 PVKRGNDTVRRKE-----ENRKLHMLKSEREKLTOQIMIKTQKELNEMQ 1092
QY 356 AHLSEDCPDVPALHTELDIAIRLVNVSNOQYGGIQLQMTKKHLE-----DTAY 402
DB 1093 AQIAEE---SQIRIELQMTLSDKSDITQLRSQALHIGLSDSSIGSGPGDAEADGGF 1148
QY 403 LVEKMRG-----QFGWSEL-----A 418
DB 1149 PESRLGWLSPVNRNNTKFGWVKYVIYVSSKKILFYDSEQDKQESNPYMWILDIDLPHV 1208
QY 419 NQAPETELIFNSIOVVPRI-----HEGNISKQDETMTDLSILSSNFTLKIPLAESAE 472
DB 1209 RPVTQDVTDAKAKEIPRIQIILYANEGESKREQE-----FPPEPVE 1251
QY 473 SSNFI---GYVVAKALQHF 488
DB 1252 KSNYICHKGHEFTPLYHF 1270

RESULT 13
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa

;; TITLE OF INVENTION: A NO. 5710022e1 Nuclear Mitotic Phosphoprotein
;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Campbell and Flores
;; STREET: 4370 La Jolla Village Drive, Suite 700
;; CITY: San Diego
;; STATE: California
;; COUNTRY: USA
;; ZIP: 92122
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/328,254
;; FILING DATE: 24-OCT-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/141,239
;; FILING DATE: 22-OCT-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Campbell, Cathryn A.
;; REGISTRATION NUMBER: 31,815
;; REFERENCE/DOCKET NUMBER: P-CJ 1191
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619) 535-9001
;; TELEFAX: (619) 535-8949
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2482 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-328-254-6

Query Match 5.0%; Score 130.5; DB 1; Length 2482;
Best Local Similarity 18.8%; Pred. No. 0.0049;
Matches 99; Conservative 85; Mismatches 165; Indels 177; Gaps 22;
QY 45 KDSHCAPTWKDKTAISENLK-----SFSEVGE-----IDADE---EVKKALTGKQ 87
DB 1502 KOLHIAEKLKRERENDSKDKVENLERELQMEENOELVILDAENSKAEVETLKTQIEE 1561
QY 88 M-----KIMMERKEKH-----TNLMSTLKKCREKEQALKLLNEV 123
DB 1562 MARSILKVFELDLVTLRSEKENITQIQEKQQLSELDKLLSFKSLLEEKQA-----EI 1616
QY 124 QHLEEEERLCRESIADSWGECRSCLENNC--MRIYTTCPQSWs-----SVKNKIERF 174
DB 1617 Q--IKEESKTAVEMQLNQKLNEVAALCGDQEIIMKATEQSLDPPPIEEHQLRNSIEKL 1674
QY 175 FRKIYQFLPFPHEDNEKDLPISEKLEED----- 203
DB 1675 RARL-----EADKKQLVQLQKLESEHDLKGVNLERELERLETARTNOEHALEA 1727
QY 204 -----AOLQMEDVFSOLTVDVNSLFNRSFNVFQMOQEFDTQFSHFISDTDL 252
DB 1728 ENSKGEVETLAKITGTMQSLRGLDVTIRSEKENITNLOKEQER-----ISELEI 1781
QY 253 TEPYFFPAPFKEPMTKADLQCCWDIPNFOLFNCNFSVYSVESVETITMKLKA--TEDLPK 311
DB 1782 INSSFENILQEKQEKQVQMK-----KSTAMEMLQTLQKELNE 1820
QY 312 QDKAPDHGGLSKMLPGQDRGLCGELDONLS---RCFKFHEKCKCOAHLSEDCPDVPA 368
DB 1821 RVAALHN-----DQEAACKAKEQNLSQVCELEL-EKAQLLQ----- 1856
QY 369 HTELDEAIRLVNVSNOQYGGIQLQMTKKHLEDTAYLVEKMRGQFQWVSELANOAPETETIF 428
DB 1857 ---LDEAKNNYIVLOSSVNGLIQ-----EVEDGKQKLEKKDEE---ISRLKNIQIQEQEIV 1906

QY 429 NSIOVPRH---EGNISQDETMWD--LSILPSSNFTKIPLE 468
Db 1907 SKLSQVEGEHOLWKEQNLEURNLTVELEQKIQVLSQSNASLODPTLE 1952

RESULT 14
US-08-533-306A-6
; Sequence 6, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533.306A
; FILING DATE: September 25, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869COB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-533-306A-6

Query Match 4.9%; Score 129; DB 2; Length 816;
Best Local Similarity 20.0%; Pred. No. 0.0012;
Matches 101; Conservative 76; Mismatches 175; Indels 154; Gaps 21;
QY 4 KAEKNEGPRSRWQLHWGDIANNSGH--MKPPLLVIIVCLLWLDKSHCAPTKWTKTAISE 61
Db 73 QGEQQTSPR-----EYVDLREAGVKYLPKAPMLNGVCVW-----KGWID----- 114
QY 62 NLKSFSEYGEIDADEE-----VKKALTGIKQKIMMERKEKHTNLM-----STLKK 108
Db 115 -LQRLDGMGCLFEDEERAQOEDALAAQAFEAARRRTREFEDRDRSHREEMENEVESVTGM 173
QY 109 CREKEQKALK-----LNEVQEHLEEEERLC-----RESLADSWGEC 145
Db 174 LNEAGKAIKADVASLSQLDQTQELLOETROKLVNSTKRLQLEERNLSLQOLDEE 233
QY 146 RSLNENCMRIYTTQPSNWSVKNKIERF-----FRKIYQFLFPFHEDNEKD 192
Db 234 MEA-KONLERHISTLNQLSDSKKLQDFASTVEALEEGKKRFQKEIENL---TQYEEK 289
QY 193 LPISKEKLEDAQLQMEDVFSQLTVDYNSLNFNSFNFRNQOQDFDQ-----TFQSHF 246
Db 290 AAAYDKLEKTKNRLOQELD---DLVVLDNQRQLVSNL-EKKQRKFDQLLAEEKNISSKY 345

QY 247 ISDTDLTEPYFFAFSPKPMTKADLEQCMWDIPNFFQFCNFVSIVSYESTITKMLKAI 306
Db 346 ADERDRAE-----AEAREKETKA-----LSARALEEAELEAK 377
QY 307 EDLPKQDKAPDHGLISKMLPGODRGLCGELDONLSRCFKFHEKCKQAHLSDCPDVP 366
Db 378 EELER-----TNKMLKAEMEDLVSSKDDVGNVHELEKSKRALETQMEEMKTQLE 427
QY 367 ALHTEL----DEAIRL-VVNSNQYGOIQLMTRKHLEDATYLVKMRGQGVWSELANO 421
Db 428 ELEDELQASEDAKRLVFN-----MQALKGQF-----ERDLQA 460
QY 422 PETEIIFNSIQVVPRIHEGNISKQDE 447
Db 461 RDEQNEKRRQLQRLHVEYTELEDE 486

RESULT 15
US-08-742-923A-6
; Sequence 6, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742.923A
; FILING DATE: No. 5869611member 1, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869DVC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-742-923A-6

Query Match 4.9%; Score 129; DB 2; Length 816;
Best Local Similarity 20.0%; Pred. No. 0.0012;
Matches 101; Conservative 76; Mismatches 175; Indels 154; Gaps 21;

QY 4 KAEKNEGPRSRWQLHWGDIANNSGH--MKPPLLVIIVCLLWLDKSHCAPTKWTKTAISE 61
Db 73 QGEQQTSPR-----EYVDLREAGVKYLPKAPMLNGVCVW-----KGWID----- 114
QY 62 NLKSFSEYGEIDADEE-----VKKALTGIKQKIMMERKEKHTNLM-----STLKK 108
Db 115 -LQRLDGMGCLFEDEERAQOEDALAAQAFEAARRRTREFEDRDRSHREEMENEVESVTGM 173

Qy	109	CREEQKPAKL-----LNEVQHELEBERLC-----RESLADSGEC	145
Db	174	LNEAEGKAIKLADKVASLSSQDQELLOEETROKLNVSFKLROLEERNSLOLDLDE	233
Qy	146	RSCLENNCKRIYTTCPQSSWSSVKNIERF-----FRKIYFLFPFHEDNEKD	192
Db	234	MEA-KONLERHISTLNQLSDSKKLQDFASTIVEALEEGKKPFQKEIENL---TQOYTEK	289
Qy	193	LPISEKLIEDAQLTQMEDVFSQLTVDVNSFNFRMQQEFDO-----TFQSHF	246
Db	290	AAAYDKLEKTKNRLQQLD---DLVVLDLNQOLVSNL-EKKQRKFDOLLAEEKNISKY	345
Qy	247	ISDSDLTEPYFPFPAASKPEMTKADLEQOWDINFFQLFCNFSVSYIESVETIRKMLKAI	306
Db	346	ADERDRAE-----AEAREKETK-----LSLRALEALEAK	377
Qy	307	EDLPKODKAPDHGGLTSKMLPCGDRGLCGELDONLSRCFKPKCKQKCOAHLSDCPDVP	366
Db	378	EELER-----TNKKLKAEMEDLVSSKDDGVKNVHELEKSKRALETOMEEMKTOLE	427
Qy	367	ALHTEL-----DEAIRL-VNVVSNQYQGILQMTKKHLEDYALVEKMRGQFGVWSLANQA	421
Db	428	ELEDELQASDAQLRLEVN-----MQALKGOF-----ERDLQA	460
Qy	422	PETEIIFNSIQVVPRIHEGNISKQDE	447
Db	461	RDPQNEEKRQOLQOLHHEYETELEDE	486

Search completed: July 2, 2002, 11:43:54
Job time: 628 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:40:16 ; Search time 32.87 seconds
(without alignments)
1447.040 Million cell updates/sec

Title: US-09-722-544a-2
Perfect score: 2632
Sequence: 1 MKIKAEKNEGPRSNWQLHW.....FTGYVAKALQHFKEHFKTW 495
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	434	16.5	449	1 A41386	clusterin precursor
2	416	15.8	451	2 I50131	clusterin - quail
3	410.5	15.6	445	2 A40018	clusterin precursor
4	406	15.4	449	2 S07714	T64 protein precursor
5	403	15.3	448	2 A40714	clusterin precursor
6	402	15.3	448	2 I56335	clusterin precursor
7	400.5	15.2	439	2 A35744	apolipoprotein J -
8	400.5	15.2	446	2 A42108	clusterin precursor
9	380	14.4	447	2 A27205	clusterin precursor
10	168	6.4	191	2 I48174	sulfated glycoprot
11	160	6.1	1738	2 T14867	interaptin - slime
12	151.5	5.8	1091	2 T34107	hypothetical prote
13	151	5.7	1526	2 T41522	myosin ii - fssio
14	146	5.5	1136	2 B70356	chromosome assembl
15	145.5	5.5	594	1 A44073	C1K1 protein - yea
16	145	5.5	1388	2 S70633	serine/threonine-s
17	144	5.5	3225	2 I52300	giantin - human
18	144	5.5	3259	1 A56539	giantin - human
19	142.5	5.4	2653	1 S28261	centromere protein
20	141.5	5.4	3433	1 S28381	urothrin - human
21	141	5.4	1005	2 A64465	hypothetical prote
22	140.5	5.3	2166	2 G70163	hypothetical prote
23	139	5.3	1955	2 T30934	myosin-like protei
24	137	5.2	1084	2 G71329	hypothetical prote
25	137	5.2	1300	2 I53799	Cg1 protein - huma
26	137	5.2	1940	1 S04090	myosin heavy chain
27	136	5.2	1132	2 T00259	hypothetical prote
28	135	5.1	1060	1 A40264	kinesin-related pr
29	134.5	5.1	1957	2 T38077	hypothetical colle

ALIGNMENTS

RESULT 1

A41386

clusterin precursor [validated] - human

N:Alternate names: apolipoprotein J; complement cytolysis inhibitor SP-40; complement TRPM-2/clusterin protein
N:Contains: clusterin alpha chain; clusterin beta chain
C:Species: Homo sapiens (man)
C>Date: 03-Apr-1992 #sequence_revision 17-Nov-1995 #text_change 08-Dec-2000
C:Accession: S43646; S04662; A41386; A35833; S34056; A57816; B37816; PL0136; R:Wong, P.; Taillefer, D.; Lakins, J.; Pineault, J.; Chader, G.; Tenniswood, M. Eur. J. Biochem. 221, 917-925, 1994
A:Title: Molecular characterization of human TRPM-2/clusterin, a gene associated with

A:Reference number: S43646; MUID:94237156

A:Accession: S43646

A:Molecule type: DNA

A:Residues: 1-449 <NON>

A:Cross-references: GB:M64722; NID:g339972; PIDN:AAB06508.1; PID:g339973

R:Kirszbaum, L.; Sharpe, J.A.; Murphy, B.; d'Apice, A.J.F.; Classon, B.; Hudson, P.; EMBO J. 8, 711-718, 1989

A:Title: Molecular cloning and characterization of the novel, human complement-associ

A:Reference number: S04662; MUID:89251601

A:Accession: S04662

A:Molecule type: mRNA

A:Residues: 1-449 <KIR>

A:Cross-references: EMBL:X14723; NID:g30250; PIDN:CAA32847.1; PID:g30251

A:Note: parts of this sequence, including the amino end of the mature protein, were c R:Jenne, D.E.; Tschopp, J. Proc. Natl. Acad. Sci. U.S.A. 86, 7123-7127, 1989

A:Title: Molecular structure and functional characterization of a human complement cy tis fluid.

A:Reference number: A41386; MUID:89386692

A:Accession: A41386

A:Molecule type: mRNA

A:Residues: 2-449 <JEN>

A:Cross-references: GB:M25915; NID:gi180619; PIDN:AAA35692.1; PID:gi180620

R:de Silva, H.V.; Harmony, J.A.K.; Stuart, W.D.; Gil, C.M.; Robbins, J. Biochemistry 29, 5380-5389, 1990

A:Title: Apolipoprotein J: structure and tissue distribution.

A:Reference number: A35833; MUID:90344779

A:Accession: A35833

A:Molecule type: mRNA

A:Residues: 34-449 <DES>

A:Cross-references: GB:J02908; NID:gi178854; PIDN:AAA51765.1; PID:gi178855

R:Ghiso, J.; Matsubara, E.; Koudinov, A.; Choi-Miura, N.H.; Tomita, M.; Wisniewski, T. Biochem. J. 293, 27-30, 1993

A:Title: Research Communication. The cerebrospinal-fluid soluble form of Alzheimer's

A:Reference number: S34056; MUID:93319521

A:Accession: S34056

A:Molecule type: protein

A:Residues: 228-240, 'X', 242-246, 23-24, 'X', 26-34, 'X', 36-38, 'X', 40-41 <GHI>

R:James, R.W.; Hochstrasser, A.C.; Borghini, I.; Martin, B.; Pometta, D.; Hochstrasse Arterioscler. Thromb. 11, 645-652, 1991

A:Title: Characterization of a human high density lipoprotein-associated protein, NAI/NA
A:Reference number: A53177; MUID:91330083
A:Accession: A53177
A:Molecule type: protein
A:Residues: 229-242; 303-304, 'M', 306-312, 'X', 314-317; 397-403 <JAM>
R:de Silva, H.V.; Stuart, W.D.; Park, Y.B.; Mao, S.J.T.; Gil, C.M.; Wetterau, J.R.; Busc
J. Biol. Chem. 265, 14292-14297, 1990
A:Title: Purification and characterization of apolipoprotein J.
A:Reference number: A37816; MUID:90354412
A:Accession: A37816
A:Molecule type: protein
A:Residues: 23-46, 'H', 48-51, 'Q' <DE3>
A:Note: amino end of the alpha chain
A:Accession: B37816
A:Molecule type: protein
A:Residues: 228-257 <DE2>
A:Note: amino end of the beta chain
R:Choi, N.H.; Mazda, T.; Tomita, M.
Mol. Immunol. 26, 835-840, 1989
A:Title: A serum protein SP40.40 modulates the formation of membrane attack complex of C
A:Reference number: PL0135; MUID:90097955
A:Accession: PL0135
A:Molecule type: protein
A:Residues: 23-37 <CHO>
A:Note: this fragment was isolated from the membrane attack complex SC5b-9
A:Accession: PL0135
A:Molecule type: protein
A:Residues: 228-242 <CH2>
A:Note: this fragment was isolated from the membrane attack complex SC5b-9
R:Hochstrasser, A.C.; James, R.W.; Martin, B.M.; Harrington, M.; Hochstrasser, D.; Pomet
Appl. Theor. Electrophor. 1, 73-76, 1988
A:Title: HDL particle associated proteins in plasma and cerebrospinal fluid: identification
A:Reference number: S07433; MUID:91265608
A:Accession: S09339
A:Molecule type: protein
A:Residues: 229-240 <HOC>
A:Accession: S07433
A:Molecule type: protein
A:Residues: 24-27, 'S', 29-33 <HO2>
R:Kirsbaum, L.; Bozas, S.E.; Walker, I.D.
FEBS Lett. 297, 70-76, 1992
A:Title: SP-40.40, a protein involved in the control of the complement pathway, possesses
A:Reference number: A56293; MUID:92201397
A:Contents: annotation; disulfide bonds, carbohydrate binding sites
R:Kuntake, S.T.; Garilli, C.T.; Lau, K.; Protter, A.A.; Naya-Vigne, J.; Kane, J.P.
Biochemistry 33, 1988-1993, 1994
A:Title: Identification of proteins associated with apolipoprotein A-I-containing lipop
A:Reference number: A54223; MUID:94162201
A:Accession: E54223
A:Molecule type: protein
A:Residues: 228-246 <KUN>
A:Experimental source: apolipoprotein A-I-containing lipoproteins, plasma
A:Accession: F54223
A:Molecule type: protein
A:Residues: 23-34, 'X', 36-37 <KU2>
A:Experimental source: apolipoprotein A-I-containing lipoproteins, plasma
R:Dank, M.; Chabot, J.G.; Mercier, C.; Benabid, A.L.; Chauvin, C.; Quirion, R.; Suh, M.
Proc. Natl. Acad. Sci. U.S.A. 86, 8577-8581, 1991
A:Title: Human gliomas and epileptic foci express high levels of a mRNA related to rat
A:Reference number: I59206; MUID:92020896
A:Accession: I59206
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 61-449 <RES>
A:Cross-references: GB:M74816; NID:9338056; PIDN:AAAG0321.1; PID:g338057
R:Duguid, J.R.; Bohmont, C.W.; Liu, N.G.; Tourtellotte, W.W.
Proc. Natl. Acad. Sci. U.S.A. 86, 7260-7264, 1989
A:Title: Changes in brain gene expression shared by scrapie and Alzheimer disease.
A:Reference number: I48174; MUID:89386721
A:Accession: I63132
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 103-168 <RE2>

A:Cross-references: GB:M26639; NID:9338070; PIDN:AAA36609.1; PID:9553644
C:Comment: This protein has been implicated in complement cascade inhibition, membran
11 as in normal brain, in tissues affected by neurodegenerative disease processes, an
C:Genetics: This protein may assist in preventing the formation of Alzheimer's disease
A:Gene: GDB:CLU; CLT
A:Cross-references: GDB:125226; OMIM:185430
A:Map position: 8p21-8p21
A:Note: appears to be a single-copy gene; alternative exon usage in 5'-untranslated r
C:Superfamily: clusterin
C:Keywords: apoptosis; complement inhibitor; extracellular protein; glycoprotein; HDL
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-227/Domain: clusterin beta chain #status experimental <BCH>
F:228-449/Product: clusterin alpha chain #status experimental <WAT>
F:86,103,145,291,354,374/Binding site: carbohydrate (Asn) (covalent) #status experime
F:102-313,121-295,129-285/Disulfide bonds: #status experimental
F:113-305,116-302/Disulfide bonds: (or 113-302,116-305) #status experimental
F:317/Binding site: carbohydrate (Asn) (covalent) #status absent
Query Match 16.5%; Score 434; DB 1; Length 449;
Best Local Similarity 25.8%; Pred. No. 9.2e-20;
Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;
QY 30 MKPPLLIVICLLWLKDSHCAPTWK-----DKTAISENLKSFSEVEIDADBEVKKALT 83
DB 1 MKKTLFLFVGLL-----TWESQVILGDTQVSDNELQENSGSKYVKNKIQNAVN 51
QY 84 GIKQIMMERKEKHTNLSTLKCRKQKALKNLNEQVQHEERLCRESLADSWG 143
DB 52 GVQIKTLTKTWEKRTLLSNLEAKKKKDALNETRESETKLKELPVCNMTMALWE 111
QY 144 ECRSLENNCMRIYT-TCQPSWSVKNKIERFERKTYQFLFPFHEDNEKDLPISEKLIIE 202
DB 112 ECKPCLKQTCMKFYARVCRSGGLVGRQLKEEFLNQSSPFYFWMNGDR-----IDSLLEN 165
QY 203 DAQLTOMEDV----FSQLTVDVNSLNFNSFVFRQMQQEPDQTFQSHFIS-DTDLTEPYF 257
DB 166 DROQTHMLDVMDHFSRASSIIDELFQDRF-----FTREPQDTY--HYLFPSLPHRRPHF 218
QY 258 FPAFSKEPTMKADLE-QCQMDIPNFQFVNFVSIVSEVSETITKMLKAIEDLPKODKAP 316
DB 219 F--FPKSRIVRSLMPSPYEPFLNFHAMFQPFLEMIHEA-----QQAAMDIFHSPAF 267
QY 317 DHGGLISKMLPGQDGLCGELDONLSCRFKFEKQKQKQAHLSDEC-----PDVPAALHTEL 372
DB 268 QHPPTFEFIREGDDRTVCREIRHNSGTCLRMKQCKDCREILSVDCSTNNPSQAKLRREL 327
QY 373 DEAIRLVNVSNOQYQQLQMTKRKLEDVAVLVEKMRGQFGWSELANOAPETELIFNSIQ 432
DB 328 DESLQVAERLTRKYNELLKSYQWKMLNTSSLLLEQLNEQFNWVSRLANLTQGEDQYYLRVT 387
QY 433 VVPRIHEGNISKQDETMMTDLSTLPSNSFTKIPLESAESSNFYGVVVAKALQHF-KEH 491
DB 388 TVAS-HTSDSDVSGTEVVVKLFDSDPITVTPVVEVSRKNPKFMEVAKELQYRKHH 446
RESULT 2
I50131
clusterin - quail
C:Species: Coturnix coturnix (quail)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999
C:Accession: I50131
R:Michel, D.; Chatelain, G.; Herault, Y.; Brun, G.
Eur. J. Biochem. 229, 215-223, 1995
A:Title: The expression of the avian clusterin gene can be driven by two alternative
A:Reference number: I50131; MUID:95262670
A:Accession: I50131
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-451 <MIC>
A:Cross-references: EMBL:X80760; NID:g520629; PIDN:CAA56733.1; PID:g520630

C;Genetics:

A;Gene: T64

A;Introns: 29/1; 78/3; 135/3; 275/1; 313/1; 388/3; 450/2

C;Superfamily: clusterin

Query Match 15.8%; Score 416; DB 2; Length 451;

Best Local Similarity 25.7%; Pred. No. 1.2e-18;

Matches 118; Conservative 93; Mismatches 180; Indels 68; Gaps 15;

Qy 63 LKSFSEVEIDADDEVKALTGKQKIMMERKEHTNLMSTIKKCREKQKALKLLNE 122

Db 27 LKQLSAAGSKYIDAENAINGVKMTLMDKTSKEHQAMLHTEETKKKEEAVKLALE 86

Qy 123 VOEHLEBERLCRESLADSGECSCLENKCMRIYT--TCQPSWSSVKNKITERFERKIYQF 181

Db 87 KEKQLAEKQEVNCTMLSLWEECKPCLKHTCMRVYSKMGSGGLVGRGLEEFNRSPPF 146

Qy 182 LFPFHEDNEKDLPISEKLEIE-----DAQLTOMEDVFSQLTVDVNSLNFNSVFMOMQ 236

Db 147 SIWNGERIDDLDRQORRRFEDLERFGLMEDGVEDIFQDSTQLYGPAFFFR----- 202

Qy 237 EFDQTFQSHFISDLDTEPY--FFPAFSKEPMTKADLEQCWDIPN-----FFQLFC 285

Db 203 -----TPPFGGPREAE--VPPQVRVHL-----VPRRLSRELHPFFQHPM 240

Qy 286 NFSVSIYVESVETITKMLKAIEDLPKQKADPH--GGLISKMLP--GQDRGLGELDONLS 342

Db 241 HGFHRLFOPLFEMTQHMLDG-----HGAWHEPLGGFATESRNFSTRMVCREIRNSA 294

Qy 343 RCFKFEKCKCOQAHLSDC----PDVPALHTELDEAIRLVNSNOOYGQILQMTKRHLE 398

Db 295 GCLRMRDCEKREILAVDCQTPVQSOLREQFEDALRLAERTTRYDILLSAFOEML 354

Qy 399 DTAYLVERKMGQFGVWSLANOAP-----ETEIIFNSIQVVPRIHEGNISKQDETMMTD 452

Db 355 NTSSLLDLQNLRFQGWVSRGLNTQGDNGFLQVTTVFSEK---TPNLEDP--SAPADTQVT- 408

Qy 453 LSLIPSSNFTLKIPLEESAESSNFTGYVAKALQHFKEH 491

Db 409 VOLFDSEPLSLTVPGDISWDDPRFMEIVAQALQHYKQN 447

RESULT 3

A40018

Clusterin precursor - dog

N;Alternate names: 80k secreted glycoprotein, renal; complement cytotoxic inhibitor

C;Species: Canis lupus familiaris (dog)

C;Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 13-Aug-1999

C;Accession: A40018

R;Hartmann, K.; Rauch, J.; Urban, J.; Parczyk, K.; Diel, P.; Pilarsky, C.; Appel, D.; Ha

J. Biol. Chem. 266, 9924-9931, 1991

A;Title: Molecular cloning of gp 80, a glycoprotein complex secreted by kidney cells in

F;227-445/Domain: clusterin alpha chain #status experimental <ACT>

F;86.103,145,277,350,370/Binding site: carbohydrate (Asn) #status predicted

A;Accession: A40018

A;Molecule type: mRNA

A;Residues: 1-445 <HAR>

A;Cross-references: GB:M55251; GB:M38757; NID:g163953; PIDN:AAA30846.1; PID:g163954

A;Note: parts of this sequence, including the amino ends of the alpha and beta chains,

C;Superfamily: clusterin

F;23-226/Domain: signal sequence #status predicted <SIG>

F;23-226/Domain: clusterin beta chain #status experimental <BCH>

F;23-226,227-445/Product: clusterin #status experimental <MAT>

F;227-445/Domain: clusterin alpha chain #status experimental <ACT>

F;86.103,145,277,350,370/Binding site: carbohydrate (Asn) #status predicted

Query Match

Best Local Similarity 23.9%; Pred. No. 2.7e-18;

Matches 117; Conservative 102; Mismatches 19; Indels 79; Gaps 15;

Qy 30 MKPPLLVFTVCLILWLKDSHCAPTWKDKTAISENLKSFSEVGEIDAD-----EEVKKALT 83

Db 1 MKKTLILLVGLLL-----TWNGRVLGQAVSDTELQEMSTEGSKYINKKLNALK 51

Qy 84 GIKQKIMMERKEHEHTNLMSTIKKCREKQKALKLLNEVQEHLEERLCRESLADSG 143

Db 52 GVKQIKTLIEQTNEERKSLLSNLEAKKKEDALNDKSETKLLKASQGVCDTMMALWE 111

Qy 144 ECRSCLENNCMRIYT--TCQPSWSSVKNKITERFERKIYQFLFPFHEDNEKDLPISEKLEIE 202

Db 112 ECKPCLKQTCMKFYARVCSGSLVGHQLEEFNLNCSPPYFWNGDR-----IDSLLEN 165

Qy 203 DAQLTOMEDVFSQLTVDVNSLNFNSVFMOMQOEFQDQTFQSHFIS-----DIDLTEPYFF 258

Db 166 DROOTHALDV-----MQDSFNRSASI-----MDLFDQDRFFTRPODTYHSP--F 209

Qy 259 PAFSKEPMTKADLEQCWDI-----PNFFQFCNFSVSIYVESVETITKMLKAIE--- 307

Db 210 SLFORPPFPNPKFRIARNIIPFRFQPLNFHDMFQFFEDMIHOA--QOAMDVNLHRIYHF 268

Qy 308 --DLPKQDRAPDHGGLISKMLPGQDRGLGELDONLSRCFKFHEKQCKQQAHLSEDC--- 362

Db 269 PIEFPED-----NETVCKEIRHNSGTGCLMKDKCEKCOEILSVDCSSN 312

Qy 363 -PDVPALHTELDEAIRLVNSNOOYGQILQMTKRHLEDTAYLVERKMGQFGVWSLANOA 421

Db 313 NPAQVOLRQELSLSQIAEKFTKLYDELQSQYQEKMENTSSLLKQLNEQFSWYSOLANLT 372

Qy 422 PETEIIFNSTQVVPRIHEGNISKQDETMMTDLSILPSSNFTLKIPLEESAESSNFTGYV 480

Db 373 QSDPPFYLVQTTVGSQTSNPNVPGFTKVV--VKLEFSDPITVMIPEAVSRNNPKFMETV 430

Qy 481 VAKALQHFKE 490

Db 431 AERKALQERYQ 440

RESULT 4

S07714

T64 protein precursor - Japanese quail

C;Species: Coturnix coturnix japonica (Japanese quail)

C;Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 13-Aug-1999

C;Accession: S07714

R;Michel, D.; Gillet, G.; Volovitch, M.; Pessac, B.; Calothy, G.; Brun, G.

Oncogene Res. 4, 127-136, 1989

A;Title: Expression of a novel gene encoding a 51.5 kD precursor protein is induced b

A;Reference number: S07714; MUID:89239492

A;Accession: S07714

A;Molecule type: mRNA

A;Residues: 1-449 <MIC>

A;Cross-references: EMBL:X15825; NID:g62594; PIDN:CAA33823.1; PID:g62595

C;Superfamily: clusterin

F;1-18/Domain: signal sequence #status predicted <SIG>

F;19-449/Product: T64 protein #status predicted <MAT>

Query Match

Best Local Similarity 25.7%; Pred. No. 5.1e-18;

Matches 117; Conservative 91; Mismatches 180; Indels 68; Gaps 15;

Qy 63 LKSFSEVEIDADDEVKALTGKQKIMMERKEHTNLMSTIKKCREKQKALKLLNE 122

Db 27 LKQLSAAGSKYIDAENAINGVKMTLMDKTSKEHQAMLHTEETKKKEEAVKLALE 86

Qy 123 VOEHLEBERLCRESLADSGECSCLENKCMRIYT--TCQPSWSSVKNKITERFERKIYQF 181

Db 87 KEKQLAEKQEVNCTMLSLWEECKPCLKHTCMRVYSKMGSGGLVGRGLEEFNRSPPF 146

Qy 182 LFPFHEDNEKDLPISEKLEIE-----DAQLTOMEDVFSQLTVDVNSLNFNSVFMOMQ 236

Db 147 SIWNGERIDDLDRQORRRFEDLERFGLMEDGVEDIFQDSTQLYGPAFFFR----- 202

Qy 237 EFDQTFQSHFISDLDTEPY--FFPAFSKEPMTKADLEQCWDIPN-----FFQLFC 285

```
Db 203 -----TPPGGFREAF-VPPVQRVHL-----VPRRLSRELHPFFOHPM 240
Qy 286 NFSVIYVESVETTKMLKAIEDLPKODKAPDH--GGGLSKMLP-GODRGLGELQDNLS 342
Db 241 HGFRHLFQPLFMTQHMLDG-----CHGAWHPLGGGFATESRNFSDRMVVCREIRNSA 294
Qy 343 RCFRHEKCKCKQAHLSDEC-----PDVPALHTDELDEAIRLVNVSNOQYGOILOMTRKHLE 398
Db 295 GCLMRDECBKCRILAVDCSQTDPVQSQRREQFEDALRAERFTRRYDILLSAFOAEML 354
Qy 399 DTAYLVKMRGQFGWSELANQAP-----ETEIFNSIQVPRRIHEGNSKODETMMTD 452
Db 355 NTSSLLDLNQRFGWVSRNLNLTQNGDGLQVTVFSK-----TPNLEDP--SAPADTQVT- 408
Qy 453 LSILPSSNFTLKIPLESAESSNFYGVVAKALQHF 488
Db 409 VQLFDSEPLSLTPGDISWDPRFMETVPEQALQHY 444

RESULT 5
A40714
clusterin precursor - mouse
N:Alternate names: sulfated glycoprotein 2
N:Contains: clusterin alpha chain; clusterin beta chain
C:Species: Mus musculus (house mouse)
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 13-Aug-1999
A:Accession: A40714; JN0699
R:French L.E.; Chonn, A.; Ducrest, D.; Baumann, B.; Belin, D.; Wohlwend, A.; Kiss, J.Z.
J. Cell Biol. 122, 1119-1130, 1993
A:Title: Murine clusterin: molecular cloning and mRNA localization of a gene associated
A:Reference number: A40714; MUID:93359508
A:Accession: A40714
A:Molecule type: mRNA
A:Residues: 1-448 <PRE>
A:Cross-references: GB:L08235; NID:g192596; PIDN:AAA37422.1; PID:g192597
R:Lee, K.H.; Ji, Y.M.; Lim, H.M.; Lee, S.C.; You, K.H.
Biochem. Biophys. Res. Commun. 194, 1175-1180, 1993
A:Title: Molecular cloning and sequencing of sulfated
A:Reference number: JN0699; MUID:93356785
A:Accession: JN0699
A:Molecule type: mRNA
A:Residues: 1-448 <DEE>
A:Note: the authors translated the codon CGC for residue 66 as Ala
C:Comment: This protein is involved in many different cellular processes other than those
C:Superfamily: clusterin
C:Keywords: apoptosis; disulfide bond; glycoprotein; spermatogenesis
F:1-21/Domain: signal sequence #status predicted <Sig>
F:22-226/Domain: clusterin beta chain #status predicted <BCH>
F:227-448/Domain: clusterin alpha chain #status predicted <MAT>
F:102,144,290,327,353,373/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 15.3%; Score 403; DB 2; Length 448;
Best Local Similarity 23.4%; Pred. No. 7.8e-18;
Matches 112; Conservative 109; Mismatches 196; Indels 62; Gaps 14;

Qy 34 LLVFIYCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDA-----DEEVKALTGKQ 87
Db 3 ILLICVALLI-----WDNGMVLGQEVSDNELQELSTQGSRYINKEIQNAVQGVKH 54
Qy 88 MKIMMERKEKHTNLMSTLKKCREEKQKALKLLNEVQEHLEEEERLCRESLADSGEGRS 147
Db 55 IKTLIEKTNARSKLLNSLEEAKKKEDALEDTROSEMKAFFPVCNETMMALWEECKP 114
Qy 148 CLNNCMRIYT--TCQPSWSSVKNKTERFRKIYQFLFFPHEDNEKDLIPSEKLEEDQAL 206
Db 115 CLKHTCMFYARVCRSGSLVQGLVEEFLNQSPPFYFWMNGDR-----IDSLLESDRQ 168
Qy 207 TQ-----MEDVFSQLVDVNSLFRNFVROMQOEFDQTFQSHFISDIDLTPYFPPF--- 259
Db 169 SQVLAMDQDSFARASGIIDTLFQDRF--FAR-----ELHDPHYFS-----PIGFPHKR 214
```

```
Qy 260 ---AFSKEPMTKADLEQC-WDIPNFFQLFCNFSVIVESVSETITTKMLKAIEDLPKQDKA 315
Db 215 PHFLYPKSLRSLMSPSHYGPPSFHNMQPFEMIHQAQAMDVOLHSPAFQFPDVD--- 272
Qy 316 PDHGLISKMLPGQ-DRGLCGELDQNLSCRFKHEKCKQAHLSDEC-----PDVPALHT 370
Db 273 -----FLREGEDDRTVCKEIRRNSTGCLKMGCKQCEKQELTSDCSTNNPAQANLRQ 324
Qy 371 ELDEAIRLVNVSNOQYGOILOMTRKHLEDTAYLVKMRGQFGWSELANQAPETEIFNS 430
Db 325 ELNDSQVAERTEYKELLQSFQSKMLNTSSLLQNDQFNWVSQNLANTQGEDKYLR 384
Qy 431 IQVPRRIHEGNSKODETMMTDLISLPSSNFTLKIPLESAESSNFYGVVAKALQHF 489
Db 385 VSTV-TTHSSDSEVPSPRVTEVVVKLFDSDPITVLPPEVSKDNPKFMDTVAEKALQEYR 442

RESULT 6
156335
apolipoprotein J - mouse
C:Species: Mus sp. (mouse)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 13-Aug-1999
A:Accession: I56335
R:Jordan-Starck, T.C.; Lund, S.D.; Witte, D.P.; Aronow, B.J.; Ley, C.A.; Stuart, W.D.
J. Lipid Res. 35, 194-210, 1994
A:Title: Mouse apolipoprotein J: characterization of a gene implicated in atherosclerosis
A:Reference number: I56335; MUID:94223204
A:Accession: I56335
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-448 <RES>
A:Cross-references: GB:S70244; NID:g546504; PIDN:AAB30623.1; PID:g546505
C:Genetics:
A:Gene: ApoJ
C:Superfamily: clusterin

Query Match 15.3%; Score 402; DB 2; Length 448;
Best Local Similarity 24.1%; Pred. No. 9.1e-18;
Matches 115; Conservative 108; Mismatches 201; Indels 54; Gaps 15;

Qy 30 MKPPLLVTIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDADEVKALTGKQ 88
Db 1 KMLLLCVAMLLIW--DNGMV---LGEQVSDNELQELSTQGSRYINKEIQNAVQGVKH 55
Qy 89 KIMMERKEKHTNLMSTLKKCREEKQKALKLLNEVQEHLEEEERLCRESLADSGEGRS 148
Db 56 KTLIEKTNARSKLLNSLEEAKKKEDALEDTROSEMKAFFPVCNETMMALWEECKPC 115
Qy 149 LENNCMRIYT--TCQPSWSSVKNKTERFRKIYQFLFFPHEDNEKDLIPSEKLEEDQAL 207
Db 116 LKHTCMKFYARVCRSGSLVQGLVEEFLNQSPPFYFWMNGDR-----IDSLLESDRQ 169
Qy 208 Q-----MEDVFSQLVDVNSLFRNFVROMQOEFDQTFQSHFISDIDLTPYFPPF--- 259
Db 170 QVLAMDQDSFARASGIIDTLFQDRF--FAR-----ELHDPHYFS-----PIGFPHKR 215
Qy 260 ---AFSKEPMTKADLEQC-WDIPNFFQLFCNFSVIVESVSETITTKMLKAIEDLPKQDKAP 316
Db 216 HFLYPKSLRSLMSPSHYGPPSFHNMQPFEMIHQAQAMDVOLHSPAFQFPDVD--- 272
Qy 317 DHGGLISKMLPGQ-DRGLCGELDQNLSCRFKHEKCKQAHLSDEC-----PDVPALHT 371
Db 273 -----FLREGEDDRTVCKEIRRNSTGCLKMGCKQCEKQELTSDCSTNNPAQANLRQ 325
Qy 372 LDEAIRLVNVSNOQYGOILOMTRKHLEDTAYLVKMRGQFGWSELANQAPETEIFNS 431
Db 326 LNDLSQVAERTEYKELLQSFQSKMLNTSSLLQNDQFNWVSQNLANTQGEDKYLR 385
Qy 432 QVPRRIHEGNSKODETMMTDLISLPSSNFTLKIPLESAESSNFYGVVAKALQHF 489
Db 386 STV-TTHSSDSEVPSPRVTEVVVKLFDSDPITVLPPEVSKDNPKFMDTVAEKALQEYR 442
```

RESULT 7
A35744
clusterin precursor - bovine
N:Alternate names: complement cytolysis inhibitor; glycoprotein III
N:Contains: clusterin alpha chain; clusterin beta chain
C:Species: Bos primigenius taurus (cattle)
C:Date: 05-Oct-1990 #sequence_revision 05-Oct-1990 #text_change 13-Aug-1999
C:Accession: A35744
R:Palmer, D.J.; Christie, D.L.
J. Biol. Chem. 265, 6617-6623, 1990
A:Title: The primary structure of glycoprotein III from bovine adrenal medullary chromaffin cells
A:Reference number: A35744; MUID:90216681
A:Accession: A35744
A:Molecule type: mRNA
A:Residues: 1-439 <PAL>
A:Cross-references: GB:J05391; NID:g163114; PIDN:AAA30554.1; PID:g163115
A:Note: parts of this sequence, including the amino ends of the alpha and beta chains, were determined from complementary DNA
C:Superfamily: clusterin
C:Keywords: apoptosis; disulfide bond; extracellular protein; glycoprotein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-221/Domain: clusterin beta chain #status experimental <BCH>
F:20-221,222-439/Product: clusterin #status experimental <MAT>
F:222-439/Domain: clusterin alpha chain #status experimental <ACH>
F:80,97,139,283,320,346,366/Binding site: carbohydriate (Asn) (covalent) #status predicted

Query Match 15.2%; Score 400.5; DB 2; Length 439;
Best Local Similarity 24.5%; Pred. No. 1.1e-17;
Matches 114; Conservative 97; Mismatches 183; Indels 71; Gaps 15;
QY 52 TWKDTATSE-NLKSFSEVEGEIDAEVKKALGTGKQKIMMERKEKEHTLMSTLKKCR 110
DB 13 SWESGALSDELQEMSTEGSKYVKNKALKEVKQIKTQIQENRERKLLLSLEAK 72
QY 111 EEKQALKLLNEVPHLEEEERCLRESLADSWGECRSLENNCMRIYT-TCQPSWSSYKN 169
DB 73 KKKEDALNDRDSENKLRASQVCNETMTALWEECKPCLKOTCMKFYARVCRSGSLGVGH 132
QY 170 KTERFERKIYQFLPFPHEDNEKDLPISEKLEEDAQLTQMEDVFSQTLVDVNSLNRSGFN 229
DB 133 QLEEFLLNOSPFPYFWINGDR-----IDSLMENDREQSHVMDV-----MEDSTRASS 179
QY 230 VFRQMOQEFDTQFSHF-----SDTLTEPY-----FF---PAFSKEPMKADLQEC 274
DB 180 I-----MDELFDQDFLRRLRQDTQYYSFSSFPKSRFARNVMPPLLEPF 232
QY 275 WPIPNFQLFNFSVSVSSETITKMLKAIETDLPKQKAPDHGGLSKMLPGQDRGLC 334
DB 233 ---NPHDVFQPF---YDHIHQAOQAMDAHLQ-----RTPYHPPTM-EFTENNDRTVC 277
QY 335 GELDQNLSCRFHEKCKCOAHLSEDC---PDVPAHLTEDEAIRLVNVSNOOYQIL 390
DB 278 KEIRHNSTGLCRMKDQCEKCEILEVDCSASNPOTLLRQQLNASLIQAEFSRLYDQLL 337
QY 391 QMTKRHLEDTAYLVEKMRQFGWVSLANOAPETE-----IFNSIQVVPRIHEGNTSK 444
DB 338 QSVQOKMLNTSALLKQLNQFTVWSOLANLTQSDQHYLVQFTVNSHNSDPSIPSG---- 393
QY 445 QDETMTDLSILPSSNFTKIPLESSESSNFIYVVAALQHF 489
DB 394 -----LTKVIVKLFNSFPITVTPQEVSPFNEMVNAEKALQOYR 433

RESULT 8
A42108
clusterin precursor - pig
N:Alternate names: complement cytolysis inhibitor; Cp40 protein
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
C:Accession: A42108; JC5535; PC4475
R:Dlemer, V.; Hoyle, M.; Baglioni, C.; Millis, A.J.
J. Biol. Chem. 267, 5257-5264, 1992

A:Title: Expression of porcine complement cytolysis inhibitor mRNA in cultured aortic
A:Reference number: A42108; MUID:92184774
A:Accession: A42108
A:Molecule type: mRNA
A:Residues: 1-446 <DIE>
A:Cross-references: GB:M84639; NID:g164408; PIDN:AAA31013.1; PID:g164409
A:Experimental source: aortic smooth muscle cells
A:Note: sequence extracted from NCBI backbone (NCBIN:87354, NCBI:87356)
R:Ogawa, S.; Ishibashi, Y.; Sakamoto, Y.; Kitamura, K.; Kubo, M.; Sakai, T.; Inoue, K.
Biochem. Biophys. Res. Commun. 234, 712-718, 1997
A:Title: The glycoproteins that occur in the colloids of senescent porcine pituitary
A:Reference number: JC5535; MUID:97318844
A:Accession: JC5535
A:Molecule type: DNA
A:Residues: 1-446 <OGA>
A:Accession: PC4475
A:Molecule type: protein
A:Residues: 58-66;68-77;229-247;249-251;408-436 <OG2>
A:Experimental source: pituitary gland
C:Superfamily: clusterin
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-227/Domain: clusterin beta chain #status predicted <BCH>
F:23-227,228-446/Product: clusterin #status predicted <MAT>
F:228-446/Domain: clusterin alpha chain #status predicted <ACH>

Query Match 15.2%; Score 400.5; DB 2; Length 446;
Best Local Similarity 24.4%; Pred. No. 1.1e-17;
Matches 118; Conservative 98; Mismatches 196; Indels 71; Gaps 16;
QY 34 LLFVIVCLLVKDSHCAPTWK-----DKTAISENLKSFSEVEGEIDAEVKKALGTGK 86
DB 4 LLLLVGLL-----TWENGPPWLGDRAISDKELQENSTEGSKYVKNKIKALKEVK 54
QY 87 QMKIMMERKEHTNLMSTLKKCRKEQKALKNVEQHELEERCLRESLADSWGECR 146
DB 55 QIKTLTIEQSNNEERKLSLSEEAARKKEDALNDRDTETKLGSGGLCNFTMALWEECK 114
QY 147 SCLNNCMRIYT-TCQPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLEEDAQ 205
DB 115 PCLAKOTCMKFYARVCRSGSLGVGHQLEFLNOSFPYFWINGDR-----IDSLMENDRQ 168
QY 206 LTOEMEDVFSQTLVDVNSLNRSGFNVRQMOQEFDTQFSHFIS-----DTDLTEPY----- 256
DB 169 QSHVMDI-----MEDSFNRASNI-----MDELFDQDFRNFPEPDTQFSFGSSHR 214
QY 257 ---FPAPFSKEPMKADLEQCDIPNPFQFCNFSVSVSSETITKMLKAIETDLPKQD 313
DB 215 GSLFNPKSRFARNIMPPLFTDL--NYHDMFQPF-----FDMHQAOQAMDAHLHRIYH- 268
QY 314 KAPDRGGLISKMLP--GQDRGLCGELDONLSRCFKFHEKCKCOAHLSEDC-----PDVPA 367
DB 269 -FPEAG-----VPNSNDRAVCKEIRHNSTGLCRMKDQCEKREILSVDCSASNSQMQ 321
QY 368 LHTELDEAIRLVNVSNOOYQILQMTKRHLEDTAYLVEKMRQFGWVSLANOAPETEII 427
DB 322 LRQELYTSLOMAEKFSKLYDQLLQSYQOKMLNTSLLKQLNBEQFSWSOLANLTQNDORY 381
QY 428 FNSIQVVPRIHEGNTSKQDETMTDLSILPSSNFTKIPLESSESSNFIYVVAALQHF 487
DB 382 YLQVTVV-NSHGSDFSVPSGLTKVVKLVFDSYPTILLIIPQEV--DPKFMETVAEALQQ 438
QY 488 FKE 490
DB 439 YRQ 441

RESULT 9
A27205
clusterin precursor - rat
N:Alternate names: SGP-2; sulfated glycoprotein 2; TRPM-2
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 13-Aug-1999

Qy	158	TTQCPWSVSNVKNIERFFRKIVQLPPFHEDNEKDLPTSEKLIBEDAQITQ-----	208
Db	401	ROGEDSARSMEEKATKEEIKKLKSQVLOQOQLEQDKRVOE---LTEQKVLESKA	457
Qy	209	-MEDVFSQITVDVNSLNFNSFNVFQMOQEF-----DOTFQ-----	243
Db	458	SVADEFTGLMSSUNSL--REEN--ROYEBETRSLQTNIRTLQDEVYQHODATEWKNR	513
Qy	244	-----SHFIS-----DTDLTEPYFPFPAFSKEPMTKADLEQCWDIPNFQFLC	288
Db	514	KAEYIEKENHRVQNASSHDADIIRLENEKQTMEEALEKAQDKQALRE-----	568
Qy	289	VSIYE-----SVSETITMKLKAIEDLPKQKADPHGGLISKMLPGDGRGLCGEL	344
Db	569	VRVWKREMTASITSDRQIOSLAE--KVDS-----LTRELESSRRRR--EQLQ	618
Qy	345	FKFHEKCKQCAHLSDCPDVPALHTELDEATRNVVSNQYGOILQMTKRKHELTAY	404
Db	619	LGSHD-----ETKAEKMKD-----LHEAQDEIKLTWQAGOLSKNETLTFE--	668
Qy	405	EKMKGQFGW----SELANAQPETEITFNISIQVVPRIHSGNISKQDETMTMTDLS	460
Db	669	ERLKAQAEKADKKYEETKVQLRAEDLQAA--QILSGNVESKFSMQKE-----	719
Qy	461	FTLKIPLEESAESSNFTGYVYVAKALQHFKEHFK	493
Db	720	--SKIEERTILDNHN-----KELEKRLRELK	743

```

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1526 <MO>
A:Cross-references: EMBL:AL049498; PTDN:CAR39901.1; GSPDB:GN00068; SPDB:SP000001
A:Experimental source: strain 972H; cosmid c645
C:Genetics:
  A:Gene: SPDB:SPCC645.05c
  A:Map position: 3
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:80-743/Domain: myosin motor domain homology <MO>

Query Match          5.7%; Score 151; DB 2; Length 1526;
Best Local Similarity 20.7%; Pred. No. 0.18;
Matches 93; Conservative 81; Mismatches 221; Indels 54; Gaps

QY 65 SFSEVEIDADEYVKALTGKQMKIMMERKEHTNLMSTLKKCKREEQKALKLNEVQ 124
      |||| : : : : || : : : : || : : : : || : : : : || : : : :
DB 895 SFSEFTQ--QENILQRESASLKQNNLESELEKTSKVETL-----LSEQNEIK 942

```

Db	943	EKLSLEER	----	DLDTKGELESIRENNATVLSKAEAFNEQCSKSLQETIYTKDAELDKLT	998
Qy	176	RKIYQFLPFPFHEDNEKOLPISEKLEIEEDAQLTQMEDYFSQLTVDVNSLFNRSNFVRMQ	235		
Db	999	KYISDYKTEIQEMRLTNQKMKNEKSIQEGSLSKRVKKLERE-NSTLLSDVSILKQOK	105		
Qy	236	QBED--QTQSHFTSDTDLTEPYFFPFAFSPKPMWKADLCQWDIPNPFQFCNFSVSIYE	293		
Db	1058	EELSVLKGQVELTINLNEKVNYLEADYKQPKLKKELESINQDKQYQLOATKNKELEA	111		
Qy	294	SVSEITIKMLKAIEDLP-KQKAPDHHGLLISKMLPGQDRIQGLGELDONL-----SRCEKFKH	348		

Db 1118 KVKECLNNIKSLTKELENKEKQNLSDASKYIELQ-----EIHENLLLVSDLENYK 1171

Qy 349 EKQKQCOAHLSEDCPDVPAHTELDIAIRLVNVSQ-----QYQILOMTRKHLEDTA 401

Db 1172 KKEYGLQIDL-EGKQVDVTFNQELSKKHRLDTFNHESLLRSQASYKEKLSLASSENKDL 1230

Qy 402 YLVEKMRGQFGWSELANOAPETELIENSIOVVPRIHE-GNISQDETMMDLSILPSSN 460

Db 1331 NKVSSLTKQVNELSPKASKVPELE-----RKITNLMEYSQLGKTFDEKRRKALIASRDN 1285

Qy 461 FTLKIPLEESAESNFICYVAKALOHFK 489

Db 1286 BELR-SLKSELESKRKLEVEYQKVLBEVK 1313

RESULT 14

B70356

chromosome assembly protein homolog - Aquifex aeolicus

C;Species: Aquifex aeolicus

C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Jun-2000

C;Accession: B70356

R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

Nature 392, 353-358, 1998

A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A;Reference number: A70300; MUID:98196666

A;Accession: B70356

A;Status: Preliminary;

A;Molecule type: DNA

A;Residues: 1-1156 <AQF>

A;Cross-references: GB:AE000699; NID:92983238; PIDN:AAC06839.1; PID:g2983243; GB:AE00069

A;Experimental source: strain VF5

C;Genetics:

A;Gene: xcpC

C;Superfamily: chromosome segregation protein SMC1

Query Match 5.5%; Score 146; DB 2; Length 1156;

Best Local Similarity 20.0%; Pred. No. 0.26;

Matches 89; Conservative 89; Mismatches 162; Indels 104; Gaps 17;

Qy 63 LKSFSEVGETD-----ADEEVKALGIIKOMKIMME-----RKEEHTNLMSTLKK 108

Db 163 IEEISGIGYERKKEALEELAEVELKIKEDILILEISNQLKRLKEEKEKLEKELQ 222

Qy 109 CREKEQALKLLNEVOEHLEERLCRE--SLADSNCEGRSCLENNCMRIYTTCPQSWSS 166

Db 223 IKRE-TEAKILLKEKLLKERERIINELSSLESLEDTITFQIENEKEL----- 271

Qy 167 VKMKIERFFRKIYQFLFPFHEDNEK---DLPISEKLI-BEDAQLTQMEDVFSOLTVDVNS 222

Db 272 --NERERLLKEVNEKIMPFKEKVGKFTAETENAEISIKERELKESENVRKNLEELINN 329

Qy 223 LFNRSNVFRQ---MQOEFDTQFQSHFISDTLTPEY-FPPAFSKEPMTKADLEQCWDIP 278

Db 330 LUSDKENLEREVGTQLQLEK-----LKEEYKSLKEVEERKLELEBEE----- 373

Qy 279 NFQQLCNF--SVSIYGSVSETITKMLKATEDLPKQKAPDHGGLISKMLPGQDRGLGEL 337

Db 374 --ERLAITDEVYKLEEBEKEKLETKLNSL-NKEQKELEQIRANLNKI-----ERI 421

Qy 338 DQNLSCFCFHEKQKCOAHLSEDCPDVPAHTELDIAIRLVNVSQOYGOILOMTRKHL 397

Db 422 KEDINKLISEREBEKIKEIKEQEIKRLKAIKKKEEBEELNLTQELNIYEKRLSEVRKL 481

Qy 398 E-----DTAVLVEKMRGQFGWVSELN-QAPE----- 423

Db 482 EEVLKEKGAIERVRSFSDVDFKDQKGVGSVELIRVKNPEHITATEVAGGGRKRFI 541

Qy 424 ----TEIIFNSIQVVPRIHEGNIS 443

Db 542 VVEDEEVAKECIIQAKRMNLRGFS 565

RESULT 15

A44073

CIK1 protein - yeast (Saccharomyces cerevisiae)

N;Alternate names: protein YM9646.11; protein YMR198w

C;Species: Saccharomyces cerevisiae

C;Date: 30-Sep-1993 #sequence_revision 08-Mar-1996 #text_change 21-Jul-2000

C;Accession: A44073; S50927

R;Page, B.D.; Snyder, M.

Genes Dev. 6, 1414-1429, 1992

A;Title: CIK1: a developmentally regulated spindle pole body-associated protein impor

A;Reference number: A44073; MUID:92354905

A;Accession: A44073

A;Molecule type: DNA

A;Residues: 1-594 <PAG>

A;Cross-references: EMBL:M96439; NID:g171221; PIDN:AAA34494.1; PID:g171222

A;Note: sequence extracted from NCBI backbone (NCBIN:110670, NCBIP:110672)

R;Pearson, D.; Bowman, S.

submitted to the EMBL Data Library, January 1995

A;Reference number: S50917

A;Accession: S50927

A;Molecule type: DNA

A;Residues: 1-594 <PEA>

A;Cross-references: EMBL:247815; NID:g642280; PIDN:CAA87820.1; PID:g642291; GSPDB:GNO

C;Genetics:

A;Gene: SGD:CIK1; MIPS:YMR198w

A;Cross-references: SGD:S0004811; MIPS:YMR198w

A;Map position: 13R

C;Superfamily: CIK1 protein

C;Keywords: coiled coil

Query Match 5.5%; Score 145.5; DB 1; Length 594;

Best Local Similarity 20.9%; Pred. No. 0.12;

Matches 84; Conservative 71; Mismatches 138; Indels 109; Gaps 17;

Qy 44 LKDSHCAPTWKOKTAISENLKSFSEVG-----EIDADEEVKKA----- 81

Db 147 LKDNQCDLQKKNVLSKKNIVSMHAVKQVEFENDLEELSNAKREWTYKLMVEENLKPDE 206

Qy 82 -LTG-IKQMKIMME---RK-----EKEHTNLMSTLKKCEEKQAL-KLLNEVOEHLE-EE 130

Db 207 RLTDENRQLKTEEEVNRKLFILQNEENECNKKELDKKFEIFKKVNDARIEDJGQ 266

Qy 131 ERLCR--ESLADSWGECRSCLENNCMRIYTTCPQSWSSYKNIERFRKIYQFLFPFHED 188

Db 267 ERLSKVLKLDQTHGELKENIK-----TC-----RDEFNDEKRIAGEAENVFHS 311

Qy 189 NEKDLPISEKLIIEEDAQLTQMEDVFSOLTVDVNSLNFNSFNVRMQOEFDTQFSHFIS 248

Db 312 ELAVPLKLLASTQALTQVQEEKKQVEGEANNMKKYVNELEKVVQEL----- 361

Qy 249 DTDLTPEYFPFAPSKPEMTKADLE-----QCWDIPNFQFCNFVS-----IYESVSET 298

Db 362 -----YTRQNLAISIEIKGYTRCFAYANERQMPDEPHINVDRCICENSGE- 408

Qy 299 ITKMLKAIEDLPKQKAPDHGGLISKMLPGQDRGLGELDQNLSCRCFKHEKQKCOAHL 358

Db 409 --KRQVDFRVVLEETHDKHRLNECIPFLE-----KYISKLNCSIIIV 451

Qy 359 SEDCCPDVPAHTELDIAIRLVNVSQOYGOILOMTRK--HLE 398

Db 452 VSQOQTAPMKTKLLKQLI-----EQGENYKMTNLILHLD 486

Search completed: July 2, 2002, 11:44:34

Job time: 258 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:44:01 ; Search time 17.51 Seconds
(without alignments)
1094.585 Million cell updates/sec

Title: US-09-722-544A-2
Perfect score: 2632
Sequence: 1 MKIRAEKNEGSRSSWQLHW.....FIGYVAKALQHEKHEFTW 495

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	434	16.5	449	1 CLUS_HUMAN	P10909 homo sapien
2	416	15.8	451	1 CLUS_COTJA	P14018 cuturnix co
3	413.5	15.7	447	1 CLUS_RABIT	O9XSC5 oryctolagus
4	410.5	15.6	445	1 CLUS_CANFA	P25473 canis famil
5	403	15.3	448	1 CLUS_MOUSE	O06890 mus musculu
6	400.5	15.2	439	1 CLUS_BOVIN	P17897 bos taurus
7	400.5	15.2	446	1 CLUS_PIG	Q29549 sus scrofa
8	380	14.4	447	1 CLUS_RAT	P05371 rattus norv
9	168	6.4	191	1 CLUS_MESAU	P14683 mesocricetu
10	151	5.7	1526	1 MY52_SCHPO	Q9US16 schizosacch
11	145.5	5.5	594	1 CIK1_YEAST	O01649 saccharomyc
12	144.5	5.5	1938	1 MYHD_HUMAN	O9UKX3 homo sapien
13	142.5	5.4	2683	1 CENE_HUMAN	Q02224 homo sapien
14	141.5	5.4	3433	1 UTR0_HUMAN	P46939 homo sapien
15	141	5.4	1005	1 RAS0_METJA	O58718 methanococc
16	137	5.2	1940	1 MYH3_HUMAN	P11055 homo sapien
17	136	5.2	2230	1 GOG4_HUMAN	O13439 homo sapien
18	135	5.1	1060	1 EG51_XENLA	P28025 xenopus lae
19	134.5	5.1	1957	1 YD86_SCHPO	Q10411 schizosacch
20	134	5.1	5430	1 ACFT_HUMAN	Q9UPN3 homo sapien
21	132.5	5.0	852	1 RAS0_THEMA	O9X1X1 thermotoga
22	131.5	5.0	1679	1 Y109_YEAST	P40457 saccharomyc
23	130.5	5.0	1940	1 MYH3_RAT	P12847 rattus norv
24	129.5	4.9	1427	1 REST_HUMAN	P30622 homo sapien
25	129.5	4.9	2077	1 TEGU_HSV6U	P52340 human herpe
26	129	4.9	539	1 MY53_HYDAT	P39922 hydra atten
27	129	4.9	1853	1 MY5A_MOUSE	O99104 mus musculu
28	129	4.9	2704	1 BPAL_HUMAN	O03001 homo sapien
29	128.5	4.9	971	1 Y228_BORBU	O51246 borrelia bu
30	128.5	4.9	1966	1 MY5B_CAEEL	P02566 caenorhabdi
31	128	4.9	1189	1 SCII_CHICK	O90988 gallus gall
32	128	4.9	1939	1 MYH1_HUMAN	P12882 homo sapien
33	128	4.9	2869	1 RBPI_PLAVB	Q00798 plasmodium

34	127.5	4.8	1978	1 MYHB_CHICK	P10587 gallus gall
35	127	4.8	793	1 KATA_ARATH	Q07970 arabidopsis
36	127	4.8	1829	1 MY5A_CHICK	Q02440 gallus gall
37	126	4.8	1941	1 MYH2_HUMAN	Q9UKX2 homo sapien
38	126	4.8	1960	1 MYH9_HUMAN	P35579 homo sapien
39	126	4.8	2418	1 SPCA_HUMAN	P02549 homo sapien
40	125.5	4.8	571	1 YH4_YEAST	P38871 saccharomyc
41	125.5	4.8	1969	1 MY5A_CAEEL	P12844 caenorhabdi
42	125.5	4.8	3210	1 CENF_HUMAN	P49454 homo sapien
43	125	4.7	978	1 RAS0_AQUAE	O67124 aquifex aeo
44	125	4.7	1790	1 US01_YEAST	P25386 saccharomyc
45	124.5	4.7	1102	1 MY5C_CHICK	P29616 gallus gall

ALIGNMENTS

RESULT	1
CLUS_HUMAN	
ID	CLUS_HUMAN STANDARD; PRT; 449 AA.
AC	P10909; P11380; P11381;
DT	01-JUL-1989 (Rel. 11, Created)
DT	01-JUL-1989 (Rel. 11, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	Clusterin precursor (Complement-associated protein SP-40,40)
DE	(Complement cytolysis inhibitor) (CLI) (NAL and NA2) (Apolipoprotein J) (Apo-J) (TRPM-2).
GN	CLU.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC	TISSUE=Liver;
RA	MEDLINE=89251601; PubMed=2721499;
RA	Kirsbaum L., Sharpe J.A., Murphy B., D'Apice J.F.A., Classon B.,
RA	Hudson P., Walker I.D.;
RT	"Molecular cloning and characterization of the novel, human
RT	complement-associated protein, SP-40,40: a link between the
RT	complement and reproductive systems.";
RL	EMBO J. 8:711-718(1989).
RN	[2]
RP	SEQUENCE OF 2-449 FROM N.A.
RX	MEDLINE=89386692; PubMed=2780565;
RA	Jenne D.E., Tschopp J.;
RT	"Molecular structure and functional characterization of a human
RT	complement cytolysis inhibitor found in blood and seminal plasma:
RT	identity to sulfated glycoprotein 2, a constituent of rat testis
RT	fluid.";
RL	Proc. Natl. Acad. Sci. U.S.A. 86:7123-7127(1989).
RN	[3]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=90344779; PubMed=1974459;
RA	de Silva H.V., Harmony J.A.K., Stuart W.D., Gil C.M., Robbins J.;
RT	"Apolipoprotein J: structure and tissue distribution.";
RL	Biochemistry 29:5380-5389(1990).
RN	[4]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=94237156; PubMed=8181474;
RA	Wong P., Taillefer D., Lalkins J., Pineault J., Chader G.,
RT	"Molecular characterization of human TRPM-2/clusterin, a gene
RT	associated with sperm maturation, apoptosis and neurodegeneration.";
RL	Eur. J. Biochem. 221:917-925(1994).
RN	[5]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=93186813; PubMed=7680346;
RA	Pineault J.M., Tenniswood M.;
RT	"Genomic organization and expression of the rat TRPM-2 (clusterin)
RT	gene, a gene implicated in apoptosis.";
RL	J. Biol. Chem. 268:5021-5031(1993).
RN	[6]

RP SEQUENCE FROM N.A.
RA Blechschmidt K., Rosenthal A.;
RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 61-449 FROM N.A.
RX MEDLINE-92020896; PubMed-1924317;
RA Danik M., Chabot J.G., Mercier C., Benabid A.L., Chauvin C.,
RL Quirion R., Suh M.;
RT "Human gliomas and epileptic foci express high levels of a mRNA
related to rat testicular sulfated glycoprotein 2, a purported marker
of cell death.";
RN Proc. Natl. Acad. Sci. U.S.A. 88:8577-8581(1991).
RN [8]
RP SEQUENCE OF 140-449 FROM N.A.
RX TISSUE=Liver;
RA Glew M.D., Kirszbaum L., Bozas S.E., Walker I.D.;
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE OF 23-33 AND 228-240.
RX MEDLINE-91265608; PubMed-3154963;
RA Hochstrasser A.-C., James R.W., Martin B.M., Harrington M.,
RL Hochstrasser D.F., Pometta D., Merrill C.R.;
RT "HDL particle associated proteins in plasma and cerebrospinal fluid:
identification and partial sequencing.";
RN Appl. Theor. Electrophor. 1:73-76(1988).
RN [10]
RP SEQUENCE OF 23-33; 229-242; 303-317 AND 397-403.
RX MEDLINE-91230083; PubMed-1903084;
RA James R.W., Hochstrasser A.-C., Borghini I., Martin B.M., Pometta D.,
RL Hochstrasser D.F.;
RT "Characterization of a human high density lipoprotein-associated
protein, NAL/NA2. Identity with SP-40,40, an inhibitor of complement-
mediated cytolysis.";
RN Arterioscler. Thromb. 11:645-652(1991).
RN [11]
RP SEQUENCE OF 23-52 AND 228-257.
RX MEDLINE-90354412; PubMed-2387851;
RA de Silva H., Stuart W.D., Park Y.B., Mao S.J.T., Gil C.M.,
RL Wetterau J.R., Busch S.J., Harmony J.A.K.;
RT "Purification and characterization of apolipoprotein J.";
RN J. Biol. Chem. 265:14292-14297(1990).
RN [12]
RP SEQUENCE OF 23-37 AND 228-242.
RX MEDLINE-90097955; PubMed-2601725;
RA Choi N.H., Mazda T., Tomita M.;
RL "A serum protein SP40,40 modulates the formation of membrane attack
complex of complement on erythrocytes.";
RN Mol. Immunol. 26:835-840(1989).
RN [13]
RP SEQUENCE OF 23-41 AND 228-246.
RX MEDLINE-93319521; PubMed-8328966;
RA Ghiso J., Matsubara E., Koudinov A., Choi-Miura N.-H., Tomita M.,
RL Wisniewski T., Frangione B.;
RT "The cerebrospinal-fluid soluble form of Alzheimer's amyloid beta is
complexed to SP-40,40 (apolipoprotein J), an inhibitor of the
complement membrane-attack complex.";
RN Biochem. J. 293:27-30(1993).
RN [14]
RP PARTIAL SEQUENCE, DISULFIDE BONDS, AND CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE-92201397; PubMed-1551440;
RA Kirszbaum L., Bozas S.E., Walker I.D.;
RL "SP-40,40, a protein involved in the control of the complement
pathway, possesses a unique array of disulphide bridges.";
RN FEBS Lett. 297:70-76(1992).
RN [15]
RP DISULFIDE BONDS.
RX MEDLINE-93147076; PubMed-1491011;
RA Choi-Miura N.H., Takahashi Y., Nakano Y., Tobe T., Tomita M.;
RL "Identification of the disulfide bonds in human plasma protein
SP-40,40 (apolipoprotein-J).";
RN J. Biochem. 112:557-561(1992).
RN [16]
RP CARBOHYDRATE-LINKAGE SITES.

RX MEDLINE-97477902; PubMed-9336835;
RA Kapron J.T., Hilliard G.M., Lakins J.N., Tenniswood M.P., West K.A.,
RL Carr S.A., Crabb J.W.;
RT "Identification and characterization of glycosylation sites in human
serum clusterin.";
RN Protein Sci. 6:2120-2133(1997).
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CELL DEATH (APOPTOSIS)
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X14723; CAA32847.1; -
DR EMBL: M25915; AAA35692.1; ALT_INIT.
DR EMBL: J02908; AAA51765.1; ALT_INIT.
DR EMBL: M74816; AAA60321.1; -
DR EMBL: M64722; AAB06508.1; -
DR EMBL: L00974; AAA60567.1; -
DR EMBL: AF311103; -; NOT_ANNOTATED_CDS.
DR EMBL: M63379; AAB06507.1; -
DR EMBL: M63376; AAB06507.1; JOINED.
DR EMBL: M63377; AAB06507.1; JOINED.
DR EMBL: M63378; AAB06507.1; JOINED.
DR EMBL: A21577; CAA01560.1; -
DR PIR: A35833; A35833.
DR PIR: A37816; A37816.
DR PIR: B37816; B37816.
DR PIR: A41386; A41386.
DR PIR: PL0135; PL0135.
DR PIR: PL0136; PL0136.
DR PIR: S04662; S04662.
DR PIR: S07433; S07433.
DR PIR: S09339; S09339.
DR PIR: S34056; S34056.
DR PIR: S43646; S43646.
DR SWISS-2DPAGE; P10909; HUMAN.
DR MIM; 185430; -
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; CLA; 1.
DR SMART; SM00030; CLD; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
KW Apoptosis; Complement pathway; Glycoprotein; Plasma; Sulfation;
KW Signal.
FT SIGNAL
FT CHAIN 1 22
FT CHAIN 23 449
FT CHAIN 23 227
FT CHAIN 228 449
FT DISULFID 102 313
FT DISULFID 113 305
FT DISULFID 116 302
FT DISULFID 121 295
FT DISULFID 129 285
FT CARBOHYD 86 86
FT CARBOHYD 103 103
FT CARBOHYD 145 145
FT CARBOHYD 291 291
FT CARBOHYD 354 354
FT CARBOHYD 374 374
FT CONFLICT 28 28
FT CONFLICT 47 47
FT CONFLICT 52 52
FT CONFLICT 52 52
FT CONFLICT 305 305
C -> M (IN REF. 9 AND 10).
Q -> H (IN REF. 11).
G -> Q (IN REF. 11).
C -> M (IN REF. 10).

```

Query Match      16.5%; Score 434; DB 1; Length 449;
Best Local Similarity 25.8%; Pred. No. 1.6e-19;
Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

Oy 30 MKPPLLVIVCLLWLKSHCAPTK-----DKTAISENLKSFSEVGEIDAEVKKALT 83
Dy 1 MKTKLLIVFVGLL-----TWESQVLDQTVSDNELQEMNOGSKYVKNKEIONAVN 51
Oy 84 GIKQMKIMMERKEHTNLMTLKKCKEOKEALKLLNEVOEHLEERLCRESLADSWG 143
Dy 52 GVQKIKTLIEKTEKRTLLNLEBEAKKKEDALNETRESETKLKEPGVCNETMMALWE 111
Oy 144 ECRSLENNCMRIYT-TCQPSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLTIE 202
Dy 112 ECKPCLKQTKMFKYARVCRSGSLVGRQLEEFNLQSSPFYFWMNGDR-----IDSLLEN 165
Oy 203 DAQLTOMEDV----FSQLTVDVNSLFNRSFNVFMQOEFTQSFHIS--DTOLTPEYF 257
Dy 166 DRQTHMLDVQDHFSSRASSIIDELFQDRF-----FTREPODTY--HYLPFSLPHRRPHF 218
Oy 258 FPAFSKEPMTKADLE-OCWDIPNFFQLFNFSVSIYSEVSETITKMLKAIEDLPKQDKAP 316
Dy 219 F--FPKSRIVRSIMFPSPYEPLENFHMFQPLEMHEA-----QQAMDIHFHSPAF 267
Oy 317 DHGLISKMLPGQDRLGGLGELDONLSRCFKPHEKCKQCAHLSDEC-----PDVPALHTEL 372
Dy 268 QHPTEFIREGDDRTVCREIRHNSGTCLRMKQDCCKREILSVDCSTNNPNSQAKLRREL 327
Oy 373 DEARLVNSVNOQYGOILQMTKRLKLEDTAYLVEKMRGFGVHVSANQAPTEIIFNSIQ 432
Dy 328 DESQVABRLTKNELLKQWMLNTSSLEQLNQFNQVNSLANLQGEQDYILRVLT 387
Oy 433 VVPRHIEGNIKQDETMMTDLISLPSSNFTLKIPILESSESSNFIYVVAKALQHF-KEH 491
Dy 388 TVAS-HTSDSDVPVSGVTEVVVKLFDSDPITVTPVEVSRKNPKFMEIVAEKALQYRKHH 446

RESULT 2
ID CLUS_CONJA STANDARD; PRT; 451 AA.
AC P14018;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-OCT-1996 (Rel. 24, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Clusterin precursor (51.5 kDa protein).
GN T64.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroretina;
RA MEDLINE=89239492; PubMed=2541393;
RX Michel D., Gillet G., Volovitch M., Pessac B., Calothy G., Brun G.;
RT "Expression of a novel gene encoding a 51.5 kD precursor protein is
induced by different retroviral oncogenes in quail neuroretinal
cells.";
RL Oncogene Res. 4:127-136(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95262670; PubMed=7744033;
RA Michel D., Chatelain G., Herault Y., Brun G.;
RT "The expression of the avian clusterin gene can be driven by two
alternative promoters with distinct regulatory elements.";
RL Eur. J. Biochem. 229:215-223(1995).
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CELL DEATH.
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER

```

```

(CY SIMILARITY).
-1- INDUCTION: BY DIFFERENT RETROVIRAL ONCOGENES.
-1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; X15825; CAA33823.1; -
EMBL; X80760; CAA56733.1; -
PIR; S07714; S07714.
InterPro; IPR000753; Clusterin.
Pfam; PF01093; Clusterin; 1.
SMART; SM00035; CLA; 1.
SMART; SM00030; CLB; 1.
PROSITE; PS00492; CLUSTERIN_1; 1.
PROSITE; PS00493; CLUSTERIN_2; 1.
Glycoprotein; Signal.
SIGNAL 1 18
CHAIN 19 451
CHAIN 19 230
CHAIN 231 451
DISULFID 98 314
DISULFID 109 306
DISULFID 112 303
DISULFID 117 296
DISULFID 125 286
CARBOHYD 99 99
CARBOHYD 141 141
CARBOHYD 278 278
CARBOHYD 355 355
CARBOHYD 375 375
CONFLICT 437 437
CONFLICT 445 451
SEQUENCE 451 AA; 51800 MW; 8CF811225C5B22F3 CRC64;
Query Match      15.8%; Score 416; DB 1; Length 451;
Best Local Similarity 25.7%; Pred. No. 1.9e-18;
Matches 118; Conservative 93; Mismatches 180; Indels 68; Gaps 15;
Oy 63 LKSPSEVGEIDADBEVKKALTGIKIMMERKEHTNLMSTLKKCKEOKEALKLLNE 122
Dy 27 LKQLSAGSKYIDAEVNAINGVKMTLMDKTSKEHQAMLHLEETKKKKEAVKLALE 86
Oy 123 VQHELEEEERLCRESLADSWGECRSLENNCMRIYT-TCQPSWSSVKNKIERFRKIYQF 181
Dy 87 KEKQAEKQEVNCTMLSLWEECKPCLKHTCMRVYSKMHSGSLVGRQLEEFNRSPPF 146
Oy 182 LFPFHEDNEKDLPISEKLTIE-----DAQLTOMEDVFSQLTVDVNSLFNRSFNVFMQOQ 236
Dy 147 SIWNGRIDDLDDLRQQRERRFEDLEERFGLMEDVEDIFQDSTQLYGFAFPFR---- 202
Oy 237 EFDQTFQSHFISDRLTEPY--FPFAFSKEPMTKADLEKQCDWIDPN-----FFOLF 285
Dy 203 -----TPPFGGFREAF-VPPQVRVHL-----VPRRLSRELHFFHFOHPM 240
Oy 286 NFSVSIYSEVSETITKMLKAIEDLPKQDKAPDH--GGLISKMLP-GODRGLCGELQNL 342
Dy 241 HGPHRLFPQPLFEMTOHMLDG-----CHGAWEHPLGCFATESRNFSTRVMVCREIRNSA 294
Oy 343 RCFKFEKCKQCAHLSDEC-----PDVPALHTELDEAIRLVNSVNOQYGOILQMTKRLK 398
Dy 295 GCLMRDECEKREILAVDCSQTDPVQSQLREQFEDALRAERFTRFDLLSAFOAEM 354
Oy 399 DTAYLVKMRGQFGVWSLANQAP-----ETEIFNSIQVVPRIHSGNISKODETMTD 452
Dy 355 NTSSLLDLQANRQFGVWSRLNLTGNDGFLQVTVFSK-----TPMLEDP--SAPADTQVT 408

```

QY 453 LSLPLSSNTFLKIPLESSESSNFYGVVAKALQHKH 491
 Db 409 VQLFDSEPLSLTPVPGDISWDDPRFMEIVAEQALQHYKON 447

RESULT 3
 CLUS_RABIT STANDARD; PRT; 447 AA.
 AC Q9XSC5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Clusterin precursor (Apolipoprotein J) (Apo-J).
 GN CLU.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAPANESE WHITE;
 RA Miyata M., Biro S., Kaieda H., Eto H., Orihara K., Kihara T.,
 RA Obata H., Matsushita N., Matsuyama T., Tei C.;
 RT "Apolipoprotein J/clusterin is induced in vascular smooth muscle cells
 by balloon injury.";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
 CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
 CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
 CC CELL DEATH (BY SIMILARITY).
 CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
 CC -!- SUBCELLULAR LOCATION: PRESENT IN CHROMAFFIN GRANULES.
 CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL; AF118852; RAD24461.1; -;
 DR InterPro; IPR000753; Clusterin.
 DR Pfam; PF01093; Clusterin; 1.
 DR SMART; SM00035; Clu; 1.
 DR SMART; SM00030; Clu; 1.
 DR PROSITE; PS00492; CLUSTERIN_1; 1.
 KW Glycoprotein; Plasma; Signal.
 FT SIGNAL 1 22 BY SIMILARITY.
 FT CHAIN 23 447 CLUSTERIN.
 FT CHAIN 23 225 BETA-CHAIN (A CHAIN) (BY SIMILARITY).
 FT CHAIN 226 447 ALPHA-CHAIN (B CHAIN) (BY SIMILARITY).
 FT DISULFID 101 311 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 112 303 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 115 300 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 120 293 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 128 283 INTERCHAIN (BY SIMILARITY).
 FT CARBOHYD 85 85 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 289 289 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 326 326 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 447 AA; 51850 MW; COCA338EEB08590A CRC64;

Query Match 15.7%; Score 413.5; DB 1; Length 447;
 Best Local Similarity 24.1%; Pred. No. 2.7e-18;
 Matches 111; Conservative 93; Mismatches 204;
 Indels 53; Gaps 11;

QY 52 TWKDKTAISENLKSFSEVGEIDA-----DEEVKALTGKQKIMMERKEHTNLMST 105

Db 13 SWERGVLGDQLVSNELQEMSTQSGKYIDREIQNAVKGVQIKTLIEKTEERKTLSSV 72
 QY 106 LKCKREEKOEALKLLNEVOEHLLEERLCRESLADSWGECRSCLENNCMRIYT-TCOPSM 164
 Db 73 LEEAKNKEDALNETRDSKTKAPFVCNEMMALWECKPCLKQTCMKFYARCSRGS 132
 QY 165 SSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLEEDAEALQTMEDVFSQLTVDVNSLF 224
 Db 133 GLVGRQLSEFLNQSSPFYFWINGDR-----IDSLLENDROQSHVLVD-----MQDSF 179
 QY 225 NRSFNVFQMQQEFQTFQSHFISDTDLTEPYFFFAFSKEPMTKADLEOCWDIP----- 278
 Db 180 NRATGI---MDELFDQDFRFTHRPQDTFYHSPFSY---FRPPFLHYAKSRSLVRNIMPLSLYG 234
 QY 279 --NFTQFCNFVSIVESVETITKMLKAIEDLPKQD---KAPDHGGLISKMLPGQDRLG 333
 Db 235 PLNFQDMQFPFFEMIHQAQAMDVHLHSPAYQTPNVEFTGGPD-----DRAV 282
 QY 334 CGELDONLSRCFKFHEKCKQCAHLSEDC-----PDVPALHTELDEAIRLVNYSNQYQGI 389
 Db 283 CKEIRHNSGCLRMKDKQCAKQCEILLSDCSANNPSONQLRQELNDSRLAEELTKRYNEL 342
 QY 390 LQMKRKHLEDYALVEKMRGQFGWSELANQAPETEIFNSIQVVPRIEHNISKQDETM 449
 Db 343 LOSYQWKMLNTSLLDQPNQFNWYSQLANLTQGPQYLYLRVSTVTS-HTSESEAPSRVT 401
 QY 450 MTDLSLPSNFTKIPLESSESSNFYGVVAKALQHKH 490
 Db 402 EVVVKLFDSDPITITPEVSRDNPFKMETVAEKALQYRK 442

RESULT 4
 CLUS_CANF STANDARD; PRT; 445 AA.
 ID AC P25473;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Clusterin precursor (Glycoprotein 80) (GP80).
 GN CLU.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91236776; PubMed-2033078;
 RA Hartmann K., Rauch J., Urban J., Parczyk K., Diel P., Pillarsky C.,
 RA Appel D., Haase W., Mann K., Weller A., Koch-Brandt C.;
 RT "Molecular cloning of gp 80, a glycoprotein complex secreted by
 RT kidney cells in vitro and in vivo. A link to the reproductive system
 RT and to the complement cascade.";
 RL J. Biol. Chem. 266:9924-9931(1991).
 CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
 CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
 CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
 CC CELL DEATH.
 CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
 CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL; M55251; AAA30846.1; -;
 DR PIR; A40018; A40018.
 DR InterPro; IPR000753; Clusterin.
 DR Pfam; PF01093; Clusterin; 1.

```

DR SMART; SM00035; Cla; 1.
DR SMART; SM00030; CLb; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Plasma; Signal.
FT SIGNAL 1 22
FT CHAIN 23 445
FT CHAIN 23 445
FT CHAIN 23 445
FT CHAIN 23 445
FT DISULFID 102 309
FT DISULFID 113 301
FT DISULFID 116 298
FT DISULFID 121 291
FT DISULFID 129 281
FT CARBOHYD 86 86
FT CARBOHYD 103 103
FT CARBOHYD 145 145
FT CARBOHYD 277 277
FT CARBOHYD 287 287
FT CARBOHYD 350 350
FT CARBOHYD 370 370
SQ SEQUENCE 445 AA; 51789 MW; 023A37266ABEF374 CRC64;

Query Match 15.6%; Score 410.5; DB 1; Length 445;
Best Local Similarity 23.9%; Pred. No. 4.1e-18;
Matches 117; Conservative 102; Mismatches 192; Indels 79; Gaps 15;

QY 30 MKPPLLIVTCVLLKDKSHCAPTKWTKTAISENLKSFSEVEIDAD-----EEVKKALT 83
DQ 1 MKKTULLLVGLLL-----TWNGRVIGDQVSDTELOEMSTEGSKYINKEIKNALK 51
QY 84 GIKQMKIMMERKEKHTNLMSTLKKREEKQBAKLLNEVOEHLEERLCRESLADSWG 143
DQ 52 GWKQIKTLIEQTNEERKSLNLEAKKKEDALNDTKDSETKLKASGVGNDTFMALWE 111
QY 144 ECRSLENNCHRIYT-TCQPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLEE 202
DQ 112 ECKPKLQTKMFKFARVCRSGSLVGHQLEBFLNQSSPFYFWMNGDR-----IDSLLEN 165
QY 203 DAQLTQMEDVFSQLTVDVNSLNFSENFVFRMQQEFQDTFQSHFIS-----DTDLTPEYFF 258
DQ 166 DRQQTALDV-----MQDSFNRASSI-----MSELFQDFRFTREPQDTYHSP--F 209
QY 259 PAFSKEPTKADLEQCMDI-----PNFFQLFCNFVSIVSEYSETITKMLKAIE--- 307
DQ 210 SLFORRPFNPKFIARNIIPFRQPLNHDMPQFPFDMHQ-AQAQMDVNLHRIPHP 268
QY 308 --DLPKQDKAPDHGLISKMLPGQDRGLCGELDONLSRCFFHEKQKQAHLSDEC--- 362
DQ 269 PTFEPEED-----NRTVCKEIRHNSTGCLKMKDQCEKQCEILSVDCSSN 312
QY 363 -PDVPALHTELDEAIRLVNSNQYQGQILQMTKRHLEDTAYLVKMRGQFGWSELANOQA 421
DQ 313 NPAQVQLRQELNSLQIAEKETKLYDELLOSQYQKMFNTSLLKQLNQEFQSWLSOLANLT 372
QY 422 PETEIIIFNSIOV-PRHEGNISKQDEPTMTDLISILPSSNFTLKIPLEAESNFIGYV 480
DQ 373 QSEDPFYQLQVTVGTSQSDSNVPVGFQV--VKLFDSDPITVMPEAVSRNNPKFMEV 430
QY 481 VAKALQHEKE 490
DQ 431 AEKALQEQYRQ 440

RESULT 5
CLUS_MOUSE
ID CLUS_MOUSE STANDARD; PRT; 448 AA.
AC Q06890;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Clusterin precursor (Sulfated glycoprotein 2) (SGP-2) (Clusterin)

```

```

DE GN (Apolipoprotein J) (Apo-J).
OS CLU OR MSGP-2 OR APOJ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=93356785; PubMed=8352774;
RA Lee K.-H., Ji Y.-M., Lim H.-M., Lee S.-C., You K.-H.;
RT "Molecular cloning and sequencing of sulfated glycoprotein-2 cDNA
RL from testis of mouse: implications of two different mRNAs of SGP-2.";
RN 2
RP SEQUENCE FROM N.A.
RX MEDLINE=93359508; PubMed=8354695;
RA French L.E., Chonn A., Ducrest B., Baumann B., Belin D., Wohlwend A.,
RA Kiss J.Z., Sappino A.P., Tschopp J., Schifferli J.A.;
RT "Murine clusterin: molecular cloning and mRNA localization of a gene
RL associated with epithelial differentiation processes during
RT embryogenesis.";
RN 3
RP SEQUENCE FROM N.A.
RX Hodgdon B.A., Min B.H., Yan H., Farris J.A., Foster D.N.,
RA Strauch A.R.;
RL Submitted (XX-1992) to the EMBL/GenBank/DBJ databases.
RN 4
RP SEQUENCE FROM N.A.
RX STRAIN=CS7BL/6 X CBA; TISSUE=Liver;
RL MEDLINE=94223204; PubMed=8169523;
RA Jordan-Starck T.C., Lund S.D., Witte D.P., Aronow B.J., Ley C.A.,
RA Stuart W.B., Swertfeger D.K., Clayton L.R., Sells S.F., Paigen B.;
RT "Mouse apolipoprotein J: characterization of a gene implicated in
RL atherosclerosis.";
RN 5
RP J. Lipid Res. 35:194-210(1994).
CC 1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A
CC VARIETY OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS,
CC MEMBRANES, AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED
CC WITH PROGRAMMED CELL DEATH.
CC 1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC 1- TISSUE SPECIFICITY: MOST ABUNDANT IN STOMACH, LIVER, BRAIN, AND
CC TESTIS, WITH INTERMEDIATE LEVELS IN HEART, OVARY, AND KIDNEY.
CC 1- PTM: EXTENSIVELY GLYCOSYLATED WITH SULFATED N-LINKED
CC CARBOHYDRATES.
CC 1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; D14077; BAA03162.1; -
CC EMBL; L08235; AAA37422.1; -
CC EMBL; L05670; AAA37284.1; -
CC EMBL; S70244; AAB30623.1; -
CC PIR; A40714; A40714.
CC PIR; JN0699; JN0699.
CC MGD; MGI:88423; Clu.
CC InterPro; IPR000753; Clusterin.
CC Pfam; PF01093; Clusterin; 1.
CC SMART; SM00035; Cla; 1.
CC SMART; SM00030; CLb; 1.
CC PROSITE; PS00492; CLUSTERIN_1; 1.
CC PROSITE; PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Sulfation; Signal.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 448 CLUSTERIN.
FT CHAIN 22 226 BETA-CHAIN (SMALL SUBUNIT)

```

```

FT CHAIN 227 447 (BY SIMILARITY)
FT ALPHA-CHAIN (LARGE SUBUNIT)
FT (BY SIMILARITY)
FT INTERCHAIN (BY SIMILARITY)
FT INTERCHAIN (BY SIMILARITY)
FT INTERCHAIN (BY SIMILARITY)
FT INTERCHAIN (BY SIMILARITY)
FT INTERCHAIN (BY SIMILARITY)
FT INTERCHAIN (BY SIMILARITY)
FT N-LINKED (GLCNAC. . .) (PROBABLE)
FT N-LINKED (GLCNAC. . .) (PROBABLE)
FT N-LINKED (GLCNAC. . .) (PROBABLE)
FT N-LINKED (GLCNAC. . .) (PROBABLE)
FT N-LINKED (GLCNAC. . .) (PROBABLE)
FT N-LINKED (GLCNAC. . .) (PROBABLE)
FT L -> M (IN REF. 4)
FT L -> TV (IN REF. 3)
FT K -> N (IN REF. 3)
FT CONFLICT 335 336
FT CONFLICT 350 350
SQ SEQUENCE 448 AA; 51655 MW; A860600A6F8D47F6 CRC64;

Query Match 15.3%; Score 403; DB 1; Length 448;
Best Local Similarity 23.4%; Pred. No. 1.2e-17;
Matches 112; Conservative 109; Mismatches 196; Indels 62; Gaps 14;

QY 34 LTVFIYVLLWLDKSHCAPTHKDKTAISENLKSEVEGEIDA-----DEEVKALTGIKO 87
DB 3 ILLLCVALLI-----WDNGMVLGEQVSDNEIQELSTQGSRYINKEIQNAVQGVKH 54
QY 88 MKTMMERKEHTNLMSTLTKCKREEQKALLLNEVQHELEERELCRSLADSGECRS 147
DB 55 IKVLEIKTNAERKSLNLSLEAAKKKEDALEDTFDSMKLKAPEVCNEMTMALWEECP 114
QY 148 CLNNCHRIYV-TCQPSWSSVKKNIERFFKRYQFLPPHEDNEKDLPISEKLEIEDAQL 206
DB 115 CLKHTCMKFYARVCRSGSLVGQOLEEFLNQSPFFYFWMGDR-----IDSLLESDRQ 168
QY 207 TQ-----MEDVFSOLTVDVNSLFNRSNVFRMQQEDFTQSFHFIISDTDLTEPYEFP-- 259
DB 169 SQVLDAMQDSFARASGIIDTFLDQFR--FAR-----ELHDPHYFS-----PIGFPHKR 214
QY 260 ---AFSKEPTKADLEQ-WDIPNFFQLFCNFSVSVESSETITMKLKAIEDLPKQOKA 315
DB 215 PHELYPKSLRVLSMSPSHVGPFSFHMFPFEMIHQAAQMDVLHSPAFQFPDQVD-- 272
QY 316 PDHGLLISMLPGQ-DRGLGELDQNLSCFKFHEKQCKQAHLSDC-----PDVPAHHT 370
DB 273 -----FLREGEDDRTVCKEIRRNSTGCLKMKGQCKEQEILSDVDCSTNNPAQANLRQ 324
QY 371 ELDEATRLVNVNQYQIQLQMTKRLKLEDYALVEKMRGQFGVYSELANQAPETEIFNS 430
DB 325 ELNDSQVABRLTEYKELLQSQSKMLNTSLEQLNDQFNWVSQLANTQGEDKYILR 384
QY 431 IQVVPRIHEGNISKQDETMTDLISLPSSNFTLKIPLEESAESSNFTGYVAKALQHFK 489
DB 385 VSTV-THTSSDSEVPSTRVTVVVKLFDSDPITVVLPEVSKDNPKFMDTVAEKALQOYR 442

RESULT 6
ID CLUS_BOVIN STANDARD; PRT; 439 AA.
AC P17697;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Clusterin precursor (Glycoprotein III) (GPIII).
GN CLU.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

```

```

RC TISSUE=Chromaffin granules;
RA MEDLINE=9021681; PubMed=1691174;
RA Palmer D.J., Christie D.L.;
RT "The primary structure of glycoprotein III from bovine adrenal
RT medullary chromaffin granules. Sequence similarity with human serum
RL protein-40.40 and rat Sertoli cell glycoprotein.";
RL J. Biol. Chem. 265:6617-6623(1990).
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC -!- SUBCELLULAR LOCATION: PRESENT IN CHROMAFFIN GRANULES.
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J05391; AAA30554.1; -
DR PIR: A35744; A35744. Clusterin.
DR InterPro: IPR000753; Clusterin.
DR Pfam: PF01093; Clusterin; 1.
DR SMART: SM00035; CLU; 1.
DR SMART: SM00030; CLU; 1.
DR PROSITE: PS00492; CLUSTERIN_1; 1.
DR PROSITE: PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Plasma; Signal.
FT SIGNAL 1 22
FT CHAIN 23 439 CLUSTERIN.
FT CHAIN 23 439 BETA-CHAIN (A CHAIN).
FT CHAIN 222 439 ALPHA-CHAIN (B CHAIN).
FT DISULFID 96 305 INTERCHAIN (BY SIMILARITY).
FT DISULFID 107 287 INTERCHAIN (BY SIMILARITY).
FT DISULFID 110 294 INTERCHAIN (BY SIMILARITY).
FT DISULFID 115 287 INTERCHAIN (BY SIMILARITY).
FT DISULFID 123 277 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 346 346 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 439 AA; 51114 MW; 36979481A61B02DE CRC64;

Query Match 15.2%; Score 400.5; DB 1; Length 439;
Best Local Similarity 24.5%; Pred. No. 1.6e-17;
Matches 114; Conservative 97; Mismatches 183; Indels 71; Gaps 15;

QY 52 TWKDKTAISE-NLKSFSEVEGEIDADEVKALTGIQMKIMMERKEKEHTNLMSTLKKCR 110
DB 13 SWESGWAISDKELQEMSTEGSKYVNEIKNALKEVQIKTQIBOTNEERKLLLSLEAK 72
QY 111 EEKQEAALKLLNEVQHELEERELCRSLADSGECRSCLNNCHRIYV-TCQPSWSSVKN 169
DB 73 KKKEDALNDRDSENKLLKASQGVNCTMTALWEECPCLKQTCMKFYARVCRSGSLGVGH 132
QY 170 KIERFFKRYQFLPPHEDNEKDLPISEKLEIEDAQLTQMEDVFSQTLTVDVNSLFNRSFN 229
DB 133 QLEEFNLQNSPPFYFWINGDR-----IDSLMENDREQSHVMDV-----MEDSFTRASS 179
QY 230 VFQMQQEFQDTQSHFI-----SDTDLTEPY-----FF--PAFSKEPTKADLEQC 274
DB 180 I-----MDLEFQDRFFLRDPDQYQYSPFSFGRSLFFNPKSFARVMVFPFLEPF 232
QY 275 WDIPNFFQLFCNFSVSVIESVSETITMKLKAIEDLPKQDKAPDHGGLISKMLPGQDRGLC 334

```

```
Db 233 ----NFHDVDFQFP-----YDMTHQQAQMAHLQ-----RTPYHPTM-EFTENNDRTVC 277
Qy 335 GELDNLSRCFKFHEKCKCOAHLSEDC-----PDVPALHTLDEAIRLVNSNOYQGIL 390
Db 278 KEIRNSGCLRMKDQCEKCEILEVDCSASNPQTLLRQQLNASQLAEKFSRLYDOLL 337
Qy 391 QMTRKHLEDYALVEKMRGQFGVSELANAPETE-----IIFNSIQVVPRIHEGNIKS 444
Db 338 QSYQOKMLNTSALLKQLNEQFTWVSQLANLTQSDQHYLVFTVNSHNSDPSIPSG----- 393
Qy 445 QDETMTDLSILPSSNFTKPILESSESSNFVGVVAKALQHEK 489
Db 394 -----LTKVIIVKLENSFFITVTVFQEVSPFMENVAEKALQOYR 433

RESULT 7
CLUS_PIG STANDARD; PRT; 446 AA.
AC Q29549;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Clusterin precursor (Complement cytolysis inhibitor) (CLI).
GN CLI.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN
RP SEQUENCE FROM N.A.
RA MEDLINE=92184774; PubMed=1544909;
RX Diemer V., Hoyle M., Baglioni C., Malls A.J.;
RT "Expression of porcine complement cytolysis inhibitor mRNA in
RT cultured aortic smooth muscle cells. Changes during differentiation
RT in vitro.";
RL J. Biol. Chem. 267:5257-5264(1992).
CC
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS IN BRAIN AND LIVER; LOWER
CC LEVELS ARE DETECTED IN OTHER TISSUES, INCLUDING THE AORTA.
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M84639; AAA31013.1; -
CC InterPro; IPR000753; Clusterin.
CC Pfam; PF01093; Clusterin; 1.
CC SMART; SM00035; CLA; 1.
CC SMART; SM00030; CLB; 1.
CC PROSITE; PS00492; CLUSTERIN_1; 1.
CC PROSITE; PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Signal.
FT SIGNAL 1 28
FT CHAIN 29 446
FT CHAIN 29 227
FT CHAIN 228 446
FT DISULFID 102 312
FT DISULFID 113 304
FT DISULFID 116 301
FT DISULFID 121 294
FT DISULFID 129 284
FT DISULFID 86 86
FT CARBOHYD 103 103
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE FROM N.A.
```

```
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 446 AA; 51774 MW; BID5B434B66B3AA CRC64;

Query Match 15.2%; Score 400.5; DB 1; Length 446;
Best Local Similarity 24.4%; Pred. No. 1.7e-17;
Matches 118; Conservative 98; Mismatches 156; Indels 71; Gaps 16;

Qy 34 LLVFIIVCLLWLKDSHCAPTWK-----DKTAISENLKSFSEVGEIDAEEVKALTGIK 86
Db 4 LLLVLGLL-----TWENGPWVLGDKAISDKELQEMSTEGSKYVKNKEIKNALKEVK 54

Qy 87 QMKIMMERKEKEHTNLMSTLAKCKREEKQEAFLKLLNEVOEHLEEBERLCRESLAUSWGCR 146
Db 55 QIKTLIEQSENEERKSLJSSLEAAKKKEDALNDRDTETKLKSGQLCNETMMALWEECK 114

Qy 147 SLENNCMRIYT-TCOPSSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLIIEDAQ 205
Db 115 PCLQOTCMKFYARVCRSGSLVGLQLEEFNLQSSPFFYFWINGDR-----IDSLMENDRQ 168

Qy 206 LTOMEDVFSQLTVDNLSLNFNSFNVFOMQOEFDTQSFHFS-----DTDLEPY----- 256
Db 169 QSHVMDI-----MEDSFNRASNI-----MDLFDQDFRFPFDTQFFSPFGSSHR 214

Qy 257 ---FFPAPSKPEMTKADLEQCDWIDPFPOLFCNFSVSYSYSETITMKLAKIEDLPKQD 313
Db 215 GSLFFNPKSRFARNIMPPLFTDL-NYHDMFQFP-----FDMTHQQAQMAHLRIPYH- 268

Qy 314 KAPDHGGLISMLP--GODRGLCGELDONLSKCFKHEKCKCOAHLSEDC-----PDVPA 367
Db 269 -FPAG-----VPENSNDRAVCKEIRHNSGCLRMKQCEKREILSVDCSASNSQWQ 321

Qy 368 LHTELDEAIRLVNSNOYQGIQLQMKRKHLEDYALVEKMRGQFGVSELANAPETIIL 427
Db 322 LRQELTSLQMAEKFSKLYDOLLOSYQOKMLNTSLLKQLNEQFSWSQLANLTONDDRY 381

Qy 428 FNSIQVVPRIHEGNIKSQDETMTDLSILPSSNFTKPILESSESSNFVGVVAKALQH 487
Db 392 YLQVTVV-NSHGSDPSVPSGLTKVVKVVKLFDSTPTITLIIPQEVSV--DPKFMETVAEEALQ 438

Qy 488 FKE 490
Db 439 YRQ 441

RESULT 8
CLUS_RAT STANDARD; PRT; 447 AA.
AC P05371;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Clusterin precursor (Sulfated glycoprotein 2) (SGP-2) (Dimeric acid
DE glycoprotein) (DAG) (Testosterone repressed prostate message-2) (TRPM-
DE 2).
GN CLO.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-45 AND 227-241.
RX MEDLINE=88000523; PubMed=3651384;
RA Collard M.W., Griswold M.D.;
RT "Biosynthesis and molecular cloning of sulfated glycoprotein 2
RT secreted by rat Sertoli cells.";
RL Biochemistry 26:3297-3303(1987).
RN
RP SEQUENCE FROM N.A.
```


RA Pineault J.M., Tenniswood M.;
 RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89149740; PubMed=2920020;
 RA Bettuzzi S., Hilpakka R.A., Gilna P., Liao S.;
 RT "Identification of an androgen-repressed mRNA in rat ventral prostate
 as coding for sulphated glycoprotein 2 by cDNA cloning and sequence
 analysis.";
 RL Biochem. J. 257:293-296(1989).
 RN [4]
 RP SEQUENCE OF 22-51 AND 227-256.
 RX MEDLINE=88326333; PubMed=3415696;
 RA Cheng C.Y., Chen C.C., Feng Z., Marshall A., Bardin C.W.;
 RT "Rat clusterin isolated from primary Sertoli cell-enriched culture
 medium is sulfated glycoprotein-2 (SGP-2).";
 RL Biochem. Biophys. Res. Commun. 155:398-404(1988).
 RN [5]
 RP CHARACTERIZATION OF TRPM-2.
 RX MEDLINE=90134121; PubMed=2299741;
 RA Bandayk M.G., Sawczuk I.S., Olsson C.A., Katz A.E., Buttyan R.;
 RT "Characterization of the products of a gene expressed during
 androgen-programmed cell death and their potential use as a marker
 of urogenital injury.";
 RL J. Urol. 143:407-413(1990).
 CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
 CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
 CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
 CC CELL DEATH.
 CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED BY CELLS UNDERGOING PROGRAMMED
 CC DEATH AS A RESULT OF THE HORMONAL STIMULI OR A TRAUMATIC INSULT.
 CC -1- PTM: EXTENSIVELY GLYCOSYLATED WITH SULFATED N-LINKED
 CC CARBOHYDRATES.
 CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M16975; AAA41273.1; -;
 DR EMBL; M64723; AAA42298.1; -;
 DR EMBL; M64733; AAA42299.1; -;
 DR EMBL; X13231; CAA31618.1; -;
 DR PIR; A27205; A27205.
 DR PIR; A31575; A31575.
 DR PIR; B31575; B31575.
 DR PIR; A45890; A45890.
 DR PIR; S18491; S18491.
 DR InterPro; IPRO00753; Clusterin.
 DR Pfam; PF01093; Clusterin; 1.
 DR SMART; SM00035; Cla; 1.
 DR SMART; SM00030; Clb; 1.
 DR PROSITE; PS00492; CLUSTERIN_1; 1.
 DR PROSITE; PS00493; CLUSTERIN_2; 1.
 KW Sulfation; Glycoprotein; Spermatogenesis; Signal;
 FT SIGNAL 1 21
 FT CHAIN 22 447 CLUSTERIN.
 FT CHAIN 22 226 BETA-CHAIN (SMALL SUBUNIT).
 FT CHAIN 227 447 ALPHA-CHAIN (LARGE SUBUNIT).
 FT DISULFID 101 312 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 112 304 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 115 301 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 120 294 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 128 284 INTERCHAIN (BY SIMILARITY).
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (PROBABLE).

FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CONFLICT 187 187 D -> H (IN REF. 1).
 SQ SEQUENCE 447 AA; 51375 MW; 9E2FA33E5E0C146E CRC64;

 Query Match 14.4%; Score 380; DB 1; Length 447;
 Best Local Similarity 23.5%; Pred. No. 2.9e-16;
 Matches 113; Conservative 103; Mismatches 197; Indels 68; Gaps 15;

 QY 34 LLVFIVCLLWLDKSHCAPTWKDKTAISENLKSFSEVEID-----ADEVKKALTGIKO 87
 DB 4 LLLCVALL-----TWNGMVLGEQEFSDNEIQELSTOGSRYVNEIIONAVGVKH 54
 QY 88 MKTMMERKEKEHTNLMSTLTKCKREEKQEAALKLLNEQVEHLEERLCRESLADSWCECRS 147
 DB 55 IKTLIEKTNAERKSLNLEAKKKEGALDDTRDSEMKLKAFPEVCNTMAMWEEKP 114
 QY 148 CLENNCMRIYT-TCQPSWSSVKNKIERFRKVIQVLFPPHEDNEKDLPISEKLIIEDAQ 206
 DB 115 CLKHTCMKFYARVCRSGSLVGRQLEEFNLQSSPFYFMNGDR-----IDSLLESDRQ 168
 QY 207 TQ-----MEDVFSQLTVDVNSLENRSFNVRQMOQEFDTQFSHFISDTDLTEPYFP 259
 DB 169 SQVLDAQDSFTRASGIIDTLFQDRF--FTHPEQDI-----HHF-----SPMGPPHXR 214
 QY 260 ---AFSKEPMTKADLQECWDIP-NFFQLFCNFSVSIVSEVSETITMKLKAIEDLPKQDKA 315
 DB 215 PHELYPKSLRSLMPLSHVGPLSPHMFQPFDFMIHQAAQAMDVLHPALQFPDVOFL 274
 QY 316 PDHGLISLMLPGQ-DRGLCGELDQNLSCRFHEKCKQCAHLSDC-----PDVPAHHT 370
 DB 275 KE-----GEDDPTVCYKEIRHNSGTCLMKMGQCEKCEITLSVDCSTNNPAQANLRQ 324
 QY 371 ELDEAIRLVNSNOQYQGLQIMTKRKHLEDTAYLVKMRGQFGWVSELAN--QAPETEILF 428
 DB 325 ELNDSLOVAERLTQQYNELHLSLOSKMLNTSSLLQNDQFTWVSQANLTGDDQYLRV 384
 QY 429 NSIQVVPRIHEGNISKQDQMTMDLSILPSSNFTLKLPLEESAESSNFYGVYAKALQHF 488
 DB 385 STVTT-----HSSDSEVPSSRVTEVVVKLFDSPTTVVLPEEVSKDNPKFMDTVAERAKQ 440
 QY 489 K 489
 DB 441 R 441

 RESULT 9
 CLUS_MESAU
 ID CLUS_MESAU STANDARD; PRT; 191 AA.
 AC P14863;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Clusterin (Sulfated glycoprotein 2) (SGP-2) (Fragment).
 GN CLU.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE OF 24-123 FROM N.A.
 RX MEDLINE=89386721; PubMed=2780570;
 RA Duguid J.R., Bohmont C., Liu N., Tourtellotte W.W.;
 RT "Changes in brain gene expression shared by scrapie and Alzheimer
 disease.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:7260-7264(1989).
 RN [2]
 RP SEQUENCE OF 1-23 AND 124-191 FROM N.A.
 RA Duguid J.R., Bohmont C., Liu N., Tourtellotte W.W.;
 RL Submitted (AUG-1989) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY


```
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
CC CELL DEATH.  
CC  
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.  
CC -!- PTM: EXTENSIVELY GLYCOSYLATED WITH SULFATED N-LINKED  
CC CARBOHYDRATES.  
CC  
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; M26640; AAA37102.1; -;  
CC InterPro: IPR000753; Clusterin.  
DR Pfam: PF01093; Clusterin; 1.  
DR SMART; SM00035; Cla; 1.  
DR PROSITE; PS00492; CLUSTERIN_1; PARTIAL.  
DR PROSITE; PS00493; CLUSTERIN_2; 1.  
KW Sulfation; Glycoprotein; Spermatogenesis.  
FT NON_TER 1  
FT CHAIN <1 >191 CLUSTERIN.  
FT CHAIN <1 >15 BETA-CHAIN (SMALL SUBUNIT).  
FT CHAIN 16 >191 ALPHA-CHAIN (LARGE SUBUNIT).  
FT CARBOHYD 79 79 N-LINKED (GLCNAC. .) (PROBABLE).  
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (PROBABLE).  
FT NON_TER 191 191  
SQ SEQUENCE 191 AA; 22099 MW; 20A9E7E66963C619 CRC64;  
  
Query Match 6.4%; Score 168; DB 1; Length 191;  
Best Local Similarity 35.1%; Pred. No. 0.00075;  
Matches 33; Conservative 22; Mismatches 35; Indels 4; Gaps 1;  
  
QY 330 DRGLGELDONLSRCFKHEKQKQAHLSDEC-----PDVPALHTELDARLVNVSNOQ 385  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
69 DRAVCKEIRHNSGTGLKMKGQCEKQCEILSVDCSANNPAQAHLRQELNDSLQVAERLTOR 128  
QY 386 YGOILQMTKRKHLEDPAVLVEKMRGQFGWVSELAN 419  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
129 YNELLSLQTKMLNTSSLEQLNEQFNWVSQAN 162  
  
RESULT 10  
MYS2_SCHPO STANDARD; PRT; 1526 AA.  
ID Q9USI6; P78969;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin type II heavy chain 1.  
GN MYO2 OR SPC645.05C  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A., AND FUNCTION.  
RC STRAIN=972;  
RX MEDLINE=98075862; PubMed=9415380;  
RA May K.M., Watts F.Z., Jones N., Hyams J.S.;  
RT "Type II myosin involved in cytokinesis in the fission yeast,  
RT Schizosaccharomyces pombe";  
RL Cell Motil. Cytoskeleton 38:385-396(1997).  
RL [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Wood V., Rajandream M.A., Barrell B.G., Rieger M.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
```

```
CC  
CC  
CC -!- FUNCTION: REQUIRED FOR CELL DIVISION. IT IS A COMPONENT OF THE  
CC CDC12 'SPOT', A STRUCTURE THOUGHT TO MARK THE SITE OF SEPTATION.  
CC MAY WORK IN CONJUNCTION WITH MYO3.  
CC  
CC -!- SUBUNIT: BINDS TO CDC4 AND RLC1.  
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
CC  
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; U75357; AAC49908.1; -;  
CC EMBL; AL049498; CAB39901.1; -;  
CC HSSP; P08799; 1MND.  
DR InterPro; IPR000048; IQ.  
DR InterPro; IPR004009; Myosin_N.  
DR InterPro; IPR001609; myosin_head.  
DR Pfam; PF00612; IQ; 1.  
DR Pfam; PF00063; myosin_head; 1.  
DR Pfam; PF02736; Myosin_N; 1.  
DR PRINTS; PR00193; MYOSINHEAVY.  
DR PRODOM; PD000355; myosin_head; 1.  
DR SMART; SM00015; IQ; 1.  
DR SMART; SM00242; MYSC; 1.  
DR PROSITE; PS00096; IQ; 1.  
KW Myosin; Actin-binding; ATP-binding; Coiled coil; Calmodulin-binding;  
KW Alkylation.  
FT DOMAIN 1 757 MYOSIN HEAD-LIKE.  
FT DOMAIN 758 787 IQ.  
FT DOMAIN 787 1244 COILED COIL (POTENTIAL).  
FT NP_BIND 170 177 ATP (POTENTIAL).  
FT DOMAIN 634 656 ACTIN-BINDING (BY SIMILARITY).  
FT DOMAIN 734 748 ACTIN-BINDING (BY SIMILARITY).  
FT MOD_RES 674 674 ALKYLATION (BY SIMILARITY).  
FT CONFLICT 1337 1337 S->R (IN REF. 1).  
SQ SEQUENCE 1526 AA; 176430 MW; D71D51D6578192BA CRC64;  
  
Query Match 5.7%; Score 151; DB 1; Length 1526;  
Best Local Similarity 20.7%; Pred. No. 0.089;  
Matches 93; Conservative 81; Mismatches 221; Indels 54; Gaps 15;  
  
QY 65 SFSEVGELDADEVKALGTGKIMMERKEKEHTNLMSTLKKREKQKALKLNEVQ 124  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
895 SFSETKQ--QNEQLQRESASLKQINNELESELEKTSKVETL-----LSEQNELK 942  
QY 125 EHLEERLCRESLADSGGCRSCLENNC-----MRIYTCQPSWSVKNK---IERFF 175  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
943 EKLSELEK----DLDTKGELESRENNATVLSKAEFNEQCKSLQETIVTKDELDKLT 998  
QY 176 RKIYQFLFPFHEDNEKDLPISEKLEIEEDAQLTQMEDVFSQLTVDVNSLENRNFVFRQM 235  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
999 KYISDYKTEIQEMRLTNQKNKESIQQEGSLSESILKRVKKLRE-NSTLSDVSILKQK 1057  
QY 236 QEPD--QTFQSHFISDITLPEYFFAFSPKADLEQCDWIPNFQFLFCNFSVIYE 293  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
1058 EELSVLKGVOELTINNLEEKVNYLEADVQKLPKLKKELESLNDKDLQYQLQATKNKELEA 1117  
QY 294 SVSETITKMLKAIEDLP-KODKAPDHGLLSKMLPGDGLGCLDQNL----SRCKFFH 348  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
1118 KVRECLANNIKSLKELENKEEKQNLSADSKYIELQ-----EIHENLLKVSLENYK 1171  
QY 349 EKKQCKOAHLSDECPDVPALHTELDARLVNVSNOQ-----QYGQILQMTKRHLESDTA 401  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
1172 KKEBGLQLDL-EGLKQVDVTFNQELSKKKHRLDTFNHESLLRQSASYKEKLSASSENKDL 1230  
QY 402 YLYEKMKGQFGWVSELANQAPETEIFNSIQVVPRIHE-GNISQDETMMDTJLSILPSSN 460  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
1231 NKVSSLTQKVNELSPKASKVPELE-----RKITNLMHEYSQLGKTFDEKRRKALIASRDN 1285
```



```
DR InterPro: IPR001609; myosin_head.  
DR Pfam: PF00612; IQ; 2.  
DR Pfam: PF00063; myosin_head; 1.  
DR Pfam: PF02736; Myosin_N; 1.  
DR Pfam: PF01576; Myosin_tail; 1.  
DR PRINTS: PR00193; MYOSINHEAVY.  
DR PRODOM: PD000355; myosin_head; 1.  
DR SMART: SM00015; IQ; 1.  
DR SMART: SM00242; MYSC; 1.  
DR PROSITE: PS50096; IQ; 1.  
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;  
KW Multigene family.  
FT DOMAIN 1 784 MYOSIN HEAD-LIKE.  
FT DOMAIN 785 814 IO.  
FT DOMAIN 843 1938 COILED COIL (POTENTIAL).  
FT NP_BIND 179 186 ATP (POTENTIAL).  
FT DOMAIN 659 681 ACTIN-BINDING (BY SIMILARITY).  
FT DOMAIN 761 775 ACTIN-BINDING (BY SIMILARITY).  
FT MOD_RES 130 130 METHYLATION (TRI-) (POTENTIAL).  
FT MOD_RES 699 699 ALKYLATION (SH-1) (POTENTIAL).  
FT MOD_RES 709 709 ALKYLATION (SH-2) (POTENTIAL).  
SQ SEQUENCE 1938 AA; 223678 MW; 1F6D006416381CD5 CRC64;  
  
Query Match 5.5%; Score 144.5; DB 1; Length 1938;  
Best Local Similarity 20.7%; Pred. No. 0.29;  
Matches 100; Conservative 93; Mismatches 206; Indels 83; Gaps 17;  
  
QY 59 ISENKASFSEVGEIDADEV-KKALTIKQKIMMERKEHEHNMSTLKKCR-----EEK 113  
DB 1150 IIRLEAS--GATSAQIEMKREAEFKMRDRLEATLQHEATAATLKKQADSVAEL 1207  
  
QY 114 QEALKLNEVQHELEERLCRESLAD--SWGECRSLNMCNRIYTTCPSSSVKNKI 171  
DB 1208 GEQIDNLRVQKLEKEKSELKNEIDDMASNIALSKSNIERTCTVDEQSEIKAD 1267  
  
QY 172 ERFRPKYQFLPPHEDNEKDLPISEKLEIEDAQLTQMEDVFSQTVVDVNSLFRNSFNVF 231  
DB 1268 EQQTQLI-----HDLNMOKARLQQTQNGELSHRVEEKESLSQTKSQALTOOLEELK 1320  
  
QY 232 ROMQOEFD-QTFQSHFIS---DFDLTEPYFFPAFSEKPTKADLEQCV-----DIPNFF 281  
DB 1321 ROMEETAKNMAHAHQSSRHDCDLR---EQYEEQKAEQALQKALSKANSEVAQWK 1376  
  
QY 282 QLFNCFVSIVSEYSETITKMLKAIEDL-PRQDKAPDHGGLISK---MLPGQDRGLACGEL 337  
DB 1377 TKYETDAIQRTEELEEAKKLAQRLOEAEEKTETANSKCSLEKTKQRLQGEVEDLMRD 1436  
  
QY 338 DONLSRC-----FKPHEKQKQCAHLSEDCPDVPAHTELDEAIRLVNVS 383  
DB 1437 ERSHTACATLDKKQNFQKVLAEWKQKLDESQAELAQAQKESRSLSTEL---FKWRNAYE 1493  
  
QY 384 QYQGI--LQWTRKHLEDATVYVEKMRGQGVWSELANOAPET-----EII 427  
DB 1494 EVDQLETLRNKNLQEE-----ISDLTEQTAETGKNLOEAETKKLVQEE 1540  
  
QY 428 FNSIQVVPRIHEGNISKQDETM---TDLSPSSNFTLKIPLSESAESSNFIGYVVAKA 484  
DB 1541 KSDLOVALEEVGSLHEESKILRVQLELSQVKSELDKRVIEKDEETEQLKRNQSQAEEA 1600  
  
QY 485 LQ 486  
DB 1601 LQ 1602
```

```
RESULT 13  
CENE_HUMAN STANDARD; PRT; 2663 AA.  
AC Q02224;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)
```

```
DE Centromeric protein E (CENP-E protein).  
GN CENPE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-93024922; PubMed-1406971;  
RA Yen T.J., Li G., Schaar B.T., Sziliak L., Cleveland D.W.;  
RT "CENP-E is a putative kinetochore motor that accumulates just before  
RT mitosis."; RL Nature 359:536-539(1992).  
RL Nature 359:536-539(1992).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE-95196755; PubMed-7889940;  
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;  
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed  
RT microtubule motor."; RL EMBO J. 14:918-926(1995).  
RL EMBO J. 14:918-926(1995).  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE-98437347; PubMed-9763420;  
RA Chan G.K.T., Schaar B.T., Yen T.J.;  
RT "Characterization of the kinetochore binding domain of CENP-E reveals  
RT interactions with the kinetochore proteins CENP-F and HUBB1.";  
RL J. Cell Biol. 143:49-63(1998).  
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE  
CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE  
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT  
CC AND/OR SPINDLE ELONGATION.  
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.  
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING  
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS  
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.  
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.  
  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
  
CC EMBL; Z15005; CAA78727.1; -.  
DR PIR; S28261; S28261.  
DR HSP; P17119; 3KAR.  
DR MIM; 117143; -.  
DR InterPro: IPR001752; kinesin.  
DR Pfam; PF00225; kinesin; 1.  
DR PRINTS; PR00380; KINESINHEAVY.  
DR SMART; SM00129; KISC; 1.  
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.  
DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.  
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;  
KW Cell cycle; Centromere.  
FT DOMAIN 1 335 KINESIN-MOTOR.  
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).  
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).  
FT NP_BIND 86 93 ATP (BY SIMILARITY).  
SQ SEQUENCE 2663 AA; 312087 MW; CEPCL3880C8C8C8 CRC64;  
  
Query Match 5.4%; Score 142.5; DB 1; Length 2663;  
Best Local Similarity 21.5%; Pred. No. 0.56;  
Matches 111; Conservative 81; Mismatches 191; Indels 133; Gaps 23;  
  
QY 54 KDKTAISENLKSFSE-----VGEIDA-----DEEVKALTGKQMK----- 89  
DB 1200 KERVKLKELQSFETDRHRLRGYIRETEATGLQTKELKIAHILKHEQETIDELRRSVS 1259  
QY 90 -----IMMERKEHEHNMSTLKKCRKEQEAAL---KLLNEVQHELEERLCRESLA- 139
```


Db 1551 KEASLSEWLSATETELVQKSTSEGLGLDLDTEISWAKNVKDL-----KKADLNTIT 1605

Qy 105 TLKCKREEKEALKLNEVQEHLEERLCRESLADSGECRSCLENNCMRIYTTCPSPW 164

Db 1606 -----ESSAALQNLIEGSEPI--LEERLC--VLNAGWSRVRTWEDWCNTL----- 1647

Qy 165 SSVKNKTERFRKTI-----YQFLPFPHEDNEKDLPISEKLTIE-----EDAQLTQMED 211

Db 1648 MNHQNLIEFDGNVAHISTWLYQEAALLDEIEKKPTSKQEIYKRLVSELDDANL-QVEN 1706

Qy 212 VFSQTLVDVNSLFRNSFNVRQMQOEDQTFQ--SHFISDTDLTEPYFFPAFSKEPMTKA 269

Db 1707 VRDQALILNARGSSRELVEPKLAELNRNFEKVSQHKSAKL-----LIAQEP----- 1755

Qy 270 DLEQCWDIPNFFQFCNFVSIVSYESVETITKMLKAI-----EDLPKQDKAPDHGGLISK 325

Db 1756 -LYQCLVTTTFETFGVPF--SDLEKLENDIENMLKFVEKHLESSEDEKMDSESAQIEEV 1812

Qy 326 LPQDGRGLGELDONLSRCFHEKFCOKCOAHLSEDCPDVPAHTEDELDAIRLVNVSNOQ 385

Db 1813 LQGEEMLHQPMDN-----KKEKIRQLL-----LHTRYNK-IAKAIPIQQRK 1855

Qy 386 YGQILOMTRKHELTAYLVKMRQFGWVSELANQAPETELIFNSIQVVPRIHGNISQK 445

Db 1856 MGQLASGRISLLPTDYLVE-----INKILLCMDVLSLNVPELNTAIYE-DFSQ 1906

Qy 446 DETM 449

Db 1907 EDSL 1910

RESULT 15

RA50_METJA

ID RA50_METJA STANDARD: PRT: 1005 AA.

AC Q58718;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DT DNA double-strand break repair rad50 AtPase.

GN RAD50 OR MJ1322.

OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;

OC Methanococcus.

OX NCBI_TaxID=2190;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;

RA MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,

RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,

RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,

RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,

RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,

RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,

RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,

RA Klenk H.-P., Fraser K.M., Smith H.O., Woese C.R., Venter J.C.;

RT "Complete genome sequence of the methanogenic archaeon, Methanococcus

RT jannaschii";

RL Science 273:1058-1073(1996).

CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The

CC rad50/mrell complex possesses single-strand endonuclease activity

CC and ATP-dependent double-strand-specific exonuclease activity.

CC Rad50 provides an ATP-dependent control of mreII by unwinding

CC and/or repositioning DNA ends into the mreII active site (By

CC similarity).

CC -!- SUBUNIT: Forms a complex with mreII (By similarity).

CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U67572; AAB99331.1; -.

DR TIGR: MJ1322; -.

DR InterPro: IPR003439; ABC_transportr.

DR InterPro: IPR001687; ATP_GTP_A.

DR InterPro: IPR003405; SMC_C.

DR InterPro: IPR003395; SMC_N.

DR Pfam: PF02483; SMC_C; 1.

DR Pfam: PF02463; SMC_N; 1.

DR DNA repair; Hydrolyase; ATP-binding; Coiled coil; Complete proteome.

FT NP_BIND 32 39 ATP (By SIMILARITY).

FT DOMAIN 158 849 COILED COIL (POTENTIAL).

SQ SEQUENCE 1005 AA; 119387 MW; 9BBB48173E788F3 CRC64;

Query Match 5.4%; Score 141; DB 1; Length 1005;

Best Local Similarity 19.8%; Pred. No. 0.22;

Matches 90; Conservative 80; Mismatches 172; Indels 112; Gaps 17;

Qy 38 IVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDAEEVKKALTGIGKQKIMMERKEK 97

Db 137 IAFSLSLKPSSEKLETVAKLIGIDEFEKCYQRMGEI--VKEYERLERIEGELNKENYK 194

Qy 98 EHTNLMSTLKKCKREEKQEAALKLLNEVQEHLEERLCRESLADSGECRSCLENNCMRIY 157

Db 195 ELKNKMSQLEKKNKLMKMEINDKLNKIKKEFEDIEKLFNE----- 233

Qy 158 TTCQPSWSSVKNKTERFRKIYQFLPFPHEDNEKDLPISE---KLIEDAQLTQMEDVFS 214

Db 234 -----WENKLLYKFTINKL-----BERKRALKLNQELKILEYDL----- 269

Qy 215 QLTVDVNSLFRNSFNVRQMQOEDQ-----TFQSHFISDTDLTEPYFFPAFSKEP 265

Db 270 NTVVEARETLNRHKDEYEKYSLVDEIRKTESRURELKSHEVEDYKLTGK-----QLE 321

Qy 266 MTKADLEQCWDIPNFFQFCNFVSIVSYESVETITKMLKAIED-LPKQDKAPDHGGLISK 324

Db 322 IIKGDIKLEKFEIN-----KSKYRDDIDNLDTLNKKIKDEIERTVETKD---LLEE 369

Qy 325 MLPQDGRGLGELDONLSRCFHEKFCOKCOAHLSEDCPDVPAHTEDELDAIRLVNVSNO 384

Db 370 L-----KNLNEEIEKIEKRIKICECKEYEEK-----YLEEEKAVYENKMLTL 412

Qy 385 YQGOILOMTR---KHELTAYLVKMRQFGWVSELANQAPETELIFNSIQV---PRIH 438

Db 413 EYITLLQEKKSIEKNINDLETRINKL-----LBETKN--IDIESIENSLAKEIEKKKVL 464

Qy 439 EGNISKQDETMMDLSILPSSNFTLKIPLESAAE 472

Db 465 E-NLQKEKIELNKKLGEINSEIKRLKKILDELKE 497

Search completed: July 2, 2002, 11:50:18

Job time: 377 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:43:26 ; Search time 53.01 seconds
(without alignments)
1615.403 Million cell updates/sec

Title: US-09-722-544A-2

Perfect score: 2632

Sequence: 1 MKIKAEKNEGPRSRWQLHW.....FICYVAKALQHKHEKTKW 495

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2464	93.6	466	4	Q15846
2	1836.5	69.8	465	6	Q95KN1
3	1230.5	46.8	338	6	Q9N1T8
4	423	16.1	449	6	Q29482
5	401.5	15.3	448	13	Q9RGPO
6	321	12.2	372	11	Q9JK98
7	266.5	10.1	295	11	Q35510
8	204	7.8	218	11	Q9BRD1
9	160	6.1	1738	5	Q76329
10	156.5	5.9	3616	13	Q9W6V0
11	151.5	5.8	1031	5	Q18082
12	149.5	5.7	840	6	Q95JRO
13	146	5.5	1156	16	Q66878
14	145	5.5	1388	6	Q28021
15	144	5.5	1305	10	Q9FJ35
16	144	5.5	3259	4	Q14789

17	143	5.4	1931	5	Q9NCF9
18	143	5.4	1931	5	Q9VRH9
19	142.5	5.4	533	6	Q95JY2
20	141	5.4	1388	4	Q75116
21	141	5.4	1388	4	Q9UON5
22	140.5	5.3	720	4	Q9H6Q7
23	140.5	5.3	2166	16	O51465
24	140	5.3	1379	11	Q62868
25	139	5.3	1330	6	Q97961
26	139	5.3	1530	4	O43241
27	139	5.3	1955	5	O61308
28	139	5.3	5458	5	Q9U459
29	138.5	5.3	1391	11	Q922J3
30	137	5.2	1084	16	O83423
31	137	5.2	1300	4	Q13999
32	136.5	5.2	709	5	Q9GRG1
33	136	5.2	1132	4	O75065
34	135.5	5.1	470	10	Q38843
35	135.5	5.1	684	3	Q07238
36	134.5	5.1	882	17	Q96YR5
37	134.5	5.1	1708	5	Q90US6
38	134.5	5.1	1723	2	Q9JMX8
39	134.5	5.1	1958	5	O96062
40	134	5.1	1356	4	Q14707
41	133.5	5.1	1956	5	Q20641
42	132.5	5.0	1410	4	Q14221
43	132	5.0	1033	4	Q9UFE1
44	132	5.0	1133	5	Q21022
45	132	5.0	1270	4	Q96JN2

ALIGNMENTS

RESULT 1

Q15846	PRELIMINARY;	PRT;	466 AA.
AC	Q15846;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	HYPOTHETICAL 54.2 KDA PROTEIN PRECURSOR.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=NEURETINA;		
RA	Shimizu A., Nishida K., Kinoshita S., Inazawa J., Okubo K.,		
RA	Matsubara K.;		
RT	"Expression profile of active genes in human retina.";		
RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; D63813; BAA09882.1; -.		
DR	InterPro; IPR000753; Clusterin.		
DR	Pfam; PF01093; Clusterin; 3.		
DR	SMART; SM00035; CLa; 1.		
DR	SMART; SM00030; CLb; 1.		
KW	Signal; Hypothetical protein. POTENTIAL.		
FT	SIGNAL 1 20		
FT	CHAIN 21 466		
SQ	SEQUENCE 466 AA; 54214 MW; 41A603D0FD6439C4 CRC64;		
Query Match	93.6%;	Score	2464; DB 4; Length 466;
Best Local Similarity	100.0%;	Pred. No.	2.3e-161;
Matches	466; Conservative	0; Mismatches	0; Indels
		0; Gaps	0;
QY	30 MKPPLLVFVCLLWLKDSHCAPTWKDKTALSENKSFSEVGEIDAEVYKALTGKQMK	89	
Db	1 MKPPLLVFVCLLWLKDSHCAPTWKDKTALSENKSFSEVGEIDAEVYKALTGKQMK	60	
QY	90 IMMERKEHTNLMSTLKKCRREEKQKALLLNEVQVHLEERLCRESLADSNQECRSCL	149	

Db	61	IMMERKEHTNLMSTLKKCKREKQEA	LKLLNEVQEHLEERLCRESLADSWGEC	RSC	L	120
QY	150	ENNCMRIYTCOPSWSSVKNKIERFRKI	YQFLPFPHEDNEKDLPISEKLI	E	A	209
Db	121	ENNCMRIYTCOPSWSSVKNKIERFRKI	YQFLPFPHEDNEKDLPISEKLI	E	A	180
QY	210	EDVFSQLTVDVNSLFRNSNFRMQOE	PDQTFQSHFTISD	T	L	269
Db	181	EDVFSQLTVDVNSLFRNSNFRMQOE	PDQTFQSHFTISD	T	L	240
QY	270	DLEQCDWIDPNFOLFNCFSVSIV	SETITKMLKAIEDLPKODKAPDHGGLIS	K	M	329
Db	241	DLEQCDWIDPNFOLFNCFSVSIV	SETITKMLKAIEDLPKODKAPDHGGLIS	K	M	300
QY	330	DRGLCGELDQNLRSRCFKFHEK	CQQAHLSEDCPDVPALHTELDE	A	I	389
Db	301	DRGLCGELDQNLRSRCFKFHEK	CQQAHLSEDCPDVPALHTELDE	A	I	360
QY	390	LQMKRKHLEDTAVLVEKMRGQ	FGWSELQANAPETIIFNSIQVPR	I	H	449
Db	361	LQMKRKHLEDTAVLVEKMRGQ	FGWSELQANAPETIIFNSIQVPR	I	H	420
QY	450	MTDLSILPSSNFTKLPILES	AESSNFYGVYAKALQHFKE	F	K	495
Db	421	MTDLSILPSSNFTKLPILES	AESSNFYGVYAKALQHFKE	F	K	466
RESULT	2					
Q95KN1	ID	Q95KN1	PRELIMINARY;	PRT;	465	AA.
AC	Q95KN1;					
DT	01-DEC-2001	(TrEMBLrel. 19, Created)				
DT	01-DEC-2001	(TrEMBLrel. 19, Last sequence update)				
DT	01-DEC-2001	(TrEMBLrel. 19, Last annotation update)				
DE	RETINAL CLUSTERIN-LIKE	PROTEIN CLU11B	SPLICED VARIANT.			
OS	Canis familiaris	(Dog).				
OC	Eukaryota;	Metazoa;	Chordata;	Cranial;	Vertebrata;	Euteleostomi;
OX	Mammalia;	Eutheria;	Carnivora;	Fissipedia;	Canidae;	Canis.
NCBI_Taxid	9615;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=RETINA;					
RX	MEDLINE=20156379;	PubMed=10675623;				
RA	Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Aguirre G.D.;					
RT	"Molecular cloning, characterization and expression of a novel retinal					
RT	clusterin-like protein cDNA.;"					
RL	Gene 243:151-160(2000).					
DR	EMBL; AF241121; AAK49030.1; -.					
SQ	SEQUENCE 465 AA; 54402 MW; BBDE1AC512D5D3F	CRC64;				
Query Match	69.8%;	Score 1836.5;	DB 6;	Length 465;		
Best Local Similarity	73.4%;	Pred. No. 2.7e-118;				
Matches 342;	Conservative 51;	Mismatches 72;	Indels 1;	Gaps		
QY	30	MKPPELLVFI	VCLLWLKDSHCAPTWKDKTAISEN	KLKSFSEVGEIDAEDEVK	KALTGKQMK	89
Db	1	MKPSLLVFTV	LLWLLKDKCHCAPTWKDKTDMH	GLKGFSEAGDIDVDEEV	KKALIGKQMK	60
QY	90	IMMERKEHTNLMSTLKKCKREKQEA	LKLLNEVQEHLEERLCRESLADSWGEC	RSC	L	149
Db	61	IMMERKEHTNLMSTLKKCKREKQEA	LKLLNEVQEHLEERLCRESLADSWGEC	RSC	L	120
QY	150	ENNCMRIYTCOPSWSSVKNKIERFRKI	YQFLPFPHEDNEKDLPISEKLI	E	A	209
Db	121	ENNCMRIYTCOPSWSSVKNKIERFRKI	YQFLPFPHEDNEKDLPISEKLI	E	A	180
QY	210	EDVFSQLTVDVNSLFRNSNFRMQOE	PDQTFQSHFTISD	T	L	269
Db	181	ENFNQLTVDVNSLFRNSNFRMQOE	PDQTFQSHFTISD	T	L	240
QY	270	DLEQCDWIDPNFOLFNCFSVSIV	SETITKMLKAIEDLPKODKAPDHGGLIS	K	M	329

[illegible]


```

Q29482
ID Q29482 PRELIMINARY; PRT; 449 AA.
AC Q29482;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CLUSTERIN PRECURSOR
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RS
SC SEQUENCE FROM N.A.
STRAIN=OLIVER; TISSUE=TESTIS;
RA Barber J.A., Farris J.A., Froedsson M.H.T., Crabo B.G., Foster D.;
RT "Nucleotide sequence of the complementary DNA encoding Equine
Clusterin.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RC
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR EMBL; L46797; AA80313.1; -.
DR InterPro; IPR00753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; Cla; 1.
DR SMART; SM00030; CLB; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 449 CLUSTERIN.
FT FT
FT SEQUENCE 449 AA; 52154 MW; 705A60504515F7C0 CRC64;
SQ
Query Match 16.1%; Score 423; DB 6; Length 449;
Best Local Similarity 24.5%; Pred. No. 2.7e-21;
Matches 117; Conservative 100; Mismatches 203; Indels 58; Gaps
QY 34 LLVFIYCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDAEVEVKALTGKQKIMME 93
Db 4 LLLVLGLLTLENGQ--VLGDKAVSDRELQEMSTQGSYINKETKNAKLGKVKIKNLE 60
QY 94 RKEKEHTNLMTLKKRKEKQKALKNLEQVHLHEERELCRLESADSGECRSCLENNC 153
Db 61 QTNEERKSLLTGLEAKKKKEGALNDTKDSEMKLKESQGVCTWTALWEECKPCLKQTC 120
QY 154 MRLYT-TCQPSWSSVYKNKTERFRKIYQLFPFHEDNEKDLPISEKLTEEDAQLTQMEDV 212
Db 121 MKFYARVCSGSLGVGHQLLEEFNLQSSPFYFWINGDR-----IDSLLENDRQQTHVLDV 174
QY 213 FSQLTVDVNSLFRNFPVROMQGFDPQFQSHFTSDTLDPYFPFPAFSKPEMTPKADL- 271
Db 175 -----MQDSFDRASSI-----MDLFDQDRFT-REPQDTYYYSFSS-SPHRRSLL 218
QY 272 -----EQWDIP-----NFFQLFCNPSVIYSVSETITKMLKAIEDLPKQD--KAP 316
Db 219 FNPKSRFARNIHFFPMYRHLNDFMFPQPFDMHQAQAMNLHLHRLPDQLPMTFSEGD 278
QY 317 DHGGLISKMLPGDGRGLGCELDQNLRSRCFKFHEKCKQCAHLSDEC----PDVPAHLTEL 372
Db 279 NH-----DRTVCKEIRHNSTGCLUKMKDQCKECCIELUSVDCSTNNPSQMOLREL 327
QY 373 DEAIRLVNSNOQYGOGLQMTKRLHLEDYALVLEKVRGQFGWVSELANOAPETEIFNSIQ 432
Db 328 NNSLQAERKFTKLYDELLQSYQEKMLNTSSLKQLNEQFSWVSQLANLTQGEDQYVLT 387
QY 433 VYPRIHEGNISKQDETMWTDLSILPSSNETLKPILEESAESSNFTGYGVVAKALQHFKE 490
Db 388 TVSS-HNSDSEVPSGLTRVVVKLFDQSYPTVTVPEVSRNNPKFMETVAEKALQSYRQ 444

```

RESULT	5
Q9YGP0	
ID	Q9YGP0 PRELIMINARY; PRT; 448 AA.
AC	Q9YGP0;
DT	01-MAY-1999 (TrEMBLrel. 10, Created)
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE	CLUSTERIN.
OS	Gallus gallus (Chicken).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC	Gallus.
OX	NCBI_taxID=9031;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=99134326; PubMed=9933595;
RA	Mahon M.G., Lindstedt K.A., Hermann M., Nimpf J., Schneider W.J.;
RT	"Multiple involvement of clusterin in chicken ovarian follicle development. Binding to two oocyte-specific members of the low density lipoprotein receptor gene family.";
RT	J. Biol. Chem. 274:4036-4044(1999)."
RL	-I- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES, AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED CELL DEATH (BY SIMILARITY).
CC	-I- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY SIMILARITY).
CC	-I- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR	ENBL; Af119370; RAD17257.1; --
DR	InterPro: IPR000753; Clusterin.
DR	Pfam: PF01093; Clusterin; 1.
DR	SMART: SM00035; CLA; 1.
DR	SMART: SM00030; CLB; 1.
DR	PROSITE: PS00492; CLUSTERIN_1; 1.
DR	PROSITE: PS00493; CLUSTERIN_2; 1.
KW	Glycoprotein.
SQ	SEQUENCE 448 AA; 51348 MW; 9982873DCAL13D27C CRC64;

Query Match	15.3%	Score	401.5;	DB	13;	Length	448;
Best Local Similarity	26.2%;	Pred.	No. 8.1e-20;				
Matches	121;	Conservative	86;	Mismatches	178;	Indels	77;
Gaps	17;						
QY	63	LKGSFSEGEIDADDEVKALTIQIKQIMMKERKEHTNLMSTLKKCKREEKQBALKLINE	122				
DB	27	LKQLSAAGSKYIDTEVENAIVGQMTYMDKTSKEHQAMLUHTLEETKRRKEEAVKLAE	86				
QY	123	VQHELEERLCRSLADSGWGCSELENNOIRYIT-TCQPSWSVSKNIKERRFKIYQF	181				
DB	87	KEQALAEQECVNETMLSLWBECKPCLKHTCMRVYSKITCHSGSLGVGRQLEELLNRSPPF	146				
QY	182	-----LPPFHEDNEK---DLPISEKLIIEEDAQLTOMEDVFSOLTYDVVNSLNFNR	227				
DB	147	SIWVNGRIDAALLDRQQRFEDEERFGIMEDG-----VEDIFQ-----DSTQLYGPA	197				
QY	228	FNVFRMQQEFDQTFQSHFISDITDLEPY---FFFAFSKEPMTKADLEQCWDIPNFFQL--	283				
DB	198	FPFPR-----TPEFGGFEAF-VPPVQVRVL-----VPPRRLLSR	231				
QY	284	----FCNFSYIYESVSETITKMLKATEDLPKQDKADPH--GGLISKMLP--GQDGLGCGE	336				
DB	232	ELHPFLQHPVGHFRLFEFTQRMIDG-----GHGAMDWHLGLGPFESERNFTDQRMVCRE	285				
QY	337	LDQNLSCRFHFHEKQCCQAHLSDC-----PDVPALHTELDEARLVNVSNQOYQIQLQ	392				
DB	286	IRNSAGCLMRDCEKCREIHAVDCSQTDPVQSOLREQFEDALRLAERTFRDYDDLISA	345				
QY	393	TRXHLEDATVLEKMRQCFGWVSELANOAPETEIFNSIQV---VPIRHEGNISKQDETM	449				
DB	346	FOAEMLNTSLLLQNLNRFQGVNLRLNLTQDGLQVTTVFSTKPNLEDP--SAPADTQ	403				

QY 450 MTDLSILPSNFTLKIPLESAESSNFIGYVYAKALQHEKH 491
DB 404 VT-VOLFDSPLSTVPGDISWDDPRFMEIVAEOALQHYKON 444

RESULT 6

Q9JK98 PRELIMINARY; PRT; 372 AA.

AC Q9JK98; 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CLUSTERIN ISOFORM 2.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BALB/C; TISSUE=UTERUS;

RA You K.H.; Jeon J.H.;

RT "Identification of Truncated SGP-2 Lacking a Signal Peptide for

RT Secretion.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DDAJ databases.

CC -1- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY

CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,

CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED

CC CELL DEATH (BY SIMILARITY).

CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY

CC SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.

DR EMBL; AF248058; AAF67185.1; .

DR InterPro; IPR000753; Clusterin.

DR Pfam; PF01093; Clusterin; 1.

DR SMART; SM00035; CLA; 1.

DR SMART; SM00030; CLB; 1.

DR PROSITE; PS00492; CLUSTERIN_1; 1.

DR PROSITE; PS00493; CLUSTERIN_2; 1.

KW Glycoprotein.

SQ SEQUENCE 372 AA; 43109 MW; A104A7E48520FEFE CRC64;

Query Match 12.2%; Score 321; DB 11; Length 372;

Best Local Similarity 23.6%; Pred. No. 2.2e-14;

Matches 93; Conservative 83; Mismatches 170; Indels 48; Gaps 12;

QY 113 KQALKLLNVEQHEERLCRESLADSWGECRCLNENCMRIYT-TCQPSWSSVKNKI 171

DB 4 REDALEDRSEKMLKAFPEVCNETMMALWEECKPCLKHTCMKFYARVCRSGSLVGQQL 63

QY 172 ERFRKIYQFLPFPHEDNEKDLPISEKLTIEDAQLTQ---MEDVFSQLTVDVNSLFNRS 227

DB 64 EEFNLSPPFFYFMNGDR-----IDSLESDRQSQVLDAMQDSFARASGIIDTLFQDR 117

QY 228 ENVFROMQOEFDOTFQSHFTSDTLTEPYFPF-----AFSKEPMTKADLEQC-WDIPNF 280

DB 118 F-FAR-----ELHDPHYFS-----PIGFPHKRPHFYLPKSLRSLMSPSHYGPSP 163

QY 281 FQLFCNFSYIYSVSTITKMLKAIEDLPKQDAPDHGGLISKMLPGQ-DRGLCGELDQ 339

DB 164 HNMFPFFMIHQAAQAMDVLHSPAQFPDQD-----FLREGEDDRTVCKEIRR 213

QY 340 NLSCFKFHEKCKQCAHLSDC-----PDVPAHLTELDEAIRLVNSNQYGGIQLQWTRK 395

DB 214 NSTGCLAMKGCCKQCEILVSDCTNNPAQANLRQELNDSLQVAERTEQYKELLSFQS 273

QY 396 HLEDATYLVKMRGQFQSWSELANAQAPETIENSIOVPRRIHEGNISKQDEMTMDLSI 455

DB 274 KMLNTSLLQLNDQFNWVSQLANTQGEDKYILRVSTV-TTHSSDSEVPSRTEVVVKL 332

QY 456 LPSNFTLKIPLESAESSNFIGYVYAKALQHEKH 489

DB 333 FDSPTITVLPVEVSKONPKFMDTVAEKALQEVYR 366

RESULT 7

O35510

ID O35510

AC O35510;

DT 01-JAN-1998 (TREMBlrel. 05, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE CLUSTERIN.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=WISTAR; TISSUE=SEMINAL VESICLE;

RX MEDLINE=98163873; PubMed=9503143;

RA Izawa M.;

RT "Identification of a transcript predicting an alternative form of

RT sulfated glycoprotein-2 (clusterin) in rat tissues.";

RL Biochem. Mol. Biol. Int. 44:9-19(1998).

CC -1- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY

CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,

CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED

CC CELL DEATH (BY SIMILARITY).

CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY

CC SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.

DR EMBL; D11339; BAA21775.1; .

DR InterPro; IPR000753; Clusterin.

DR Pfam; PF01093; Clusterin; 1.

DR SMART; SM00035; CLA; 1.

DR SMART; SM00030; CLB; 1.

DR PROSITE; PS00492; CLUSTERIN_1; 1.

DR PROSITE; PS00493; CLUSTERIN_2; 1.

KW Glycoprotein.

SQ SEQUENCE 295 AA; 34115 MW; C1E07FA745DD5470 CRC64;

Query Match 10.1%; Score 266.5; DB 11; Length 295;

Best Local Similarity 25.2%; Pred. No. 9.1e-11;

Matches 80; Conservative 59; Mismatches 131; Indels 47; Gaps 11;

QY 133 LCRESLADSWGECRCLNENCMRIYT-TCQPSWSSVKNKIERFRKIYQFLPFPHEDNEK 191

DB 9 VCNETHMALWEECKPCLKHTCMKFYARVCRSGSLVGROLEEFNLSQSPFFYFMNGDR-- 66

QY 192 DLPISKELTIEDAQLTQ---MEDVFSQLTVDVNSLFNRSFNVRMOMQOEFDOTFOSHEI 247

DB 67 ----IDSLESDRQSQVLDAMQDSFARASGIIDTLFQDRF--FTHEPQDI-----HHF- 114

QY 248 SDTDLTEPYFPF-----AFSKEPMTKADLEQCWDIP-NFFQLFCNFSYIYSVSETIT 300

DB 115 -----SPMGFPHKRPHFYLPKSLRSLMPLSHYGLPSFHNNFQFPFDFMIHQAAQAMDV 168

QY 301 KMLKATEDLPKQDAPDHGGLISKMLPGQ-DRGLCGELDONLSRCFKFHEKCKQCAHLS 359

DB 169 QLHSPALQFPDQDVLKE-----GEDDPTVCKEIRHNSTGCLUKMGQCEKCEILS 218

QY 360 EDC-----PDVPAHLTELDEAIRLVNSNQYGGIQLQWTRKHLDEATYLVKMRGQFQGWYS 415

DB 219 VDCSTNNPAQANLRQELNDSLQVAERLTQYNELLHSLOSKMLNTSLLQLNDQFSWVS 278

QY 416 ELANQAPETETIIPNSTQ 432

DB 279 QLANLTQGDQDQYFGSPQ 295

RESULT 8

Q9ERD1

ID Q9ERD1

AC Q9ERD1;

PRELIMINARY; PRT; 218 AA.

DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CLUSTERIN (FRAGMENT)
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE DAWLEY; TISSUE=THYMUS;
 RA Park J.H., Park J.S., Ju S.K., Na S.Y., You K.H.;
 RT "Determination of clusterin mRNA expression of apoptosis induced rat
 thymocytes in vivo and in vitro."
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
 OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
 AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
 CELL DEATH (BY SIMILARITY).
 CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
 SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
 DR EMBL; AF314657; AAG31162.1; -.
 DR InterPro; IPR00753; Clusterin.
 DR Pfam; PF01093; Clusterin; 2.
 DR SMART; SM00035; CLA; 1.
 DR PROSITE; PS00493; CLUSTERIN_2; 1.
 KW Glycoprotein.
 FT NON_TER 1 1
 SQ SEQUENCE 218 AA; 24931 MW; ED2FC4425A510589 CRC64;

Query Match 7.8%; Score 204; DB 11; Length 218;
 Best Local Similarity 28.3%; Pred. No. 1.2e-06;
 Matches 47; Conservative 36; Mismatches 73; Indels 10; Gaps 3;

QY 330 DRGLGGLDQNLSCRFHEKCKQAHLSDC-----PDVPAHTEDEAIRLVNVSNOQ 385
 DB 51 DRTVCKEIRHNSGCLKMKGCKQCEILSVDCSTNNPAQANLROELNLSQVAERLTQQ 110
 QY 386 YGOILQMTKRHLEDTAVLVEKMRQFGWVSELAN--QAPETIIFNSIQVVPRIHEGNIS 443
 DB 111 YNELLHSLSQKMLNTSLLEQNDQFSWVSQLANTOGDDQYLRSIVTT-----HSSDSE 166
 QY 444 QDETMTDLTSLPSSNFTLKIPLEESAESSNFTGYVVAKALQHFK 489
 DB 167 VPSRVTEVVKLFDSPTITVVLPEVSKDNPKFMDTVAERKALQERY 212

RESULT 9
 O76329 PRELIMINARY; PRT; 1738 AA.
 ID O76329
 AC O76329
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE INTERAPTIN.
 GN ABPD.
 OS Dictyostellium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=98365466; PubMed=9700162;
 RA Rivero F.J., Kuspa A., Brokamp R., Matzner M., Noegel A.A.;
 RT "Interaptin, an actin-binding protein of the alpha-actinin superfamily
 in Dictyostellium discoideum, is developmentally and CAMP-regulated and
 associates with intracellular membrane compartments."
 RL J. Cell Biol. 142:735-750(1998).
 DR EMBL; AF057019; AAC34562.1; -.
 DR HSSP; Q01082; IBKR.
 DR InterPro; IPR001589; Actinin_act_bind.
 DR InterPro; IPR001715; Calponin_hom.

DR InterPro; IPR001990; Granin.
 DR InterPro; IPR001451; Hexapep_transf.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF00307; CH; 2.
 DR SMART; SM00033; CH; 2.
 DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
 DR PROSITE; PS00020; ACTININ_2; 1.
 DR PROSITE; PS50021; CH; 2.
 DR PROSITE; PS00422; GRANINS_1; UNKNOWN_1.
 DR PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
 SQ SEQUENCE 1738 AA; 204427 MW; 577A99D2EC79AF5C CRC64;

Query Match 6.1%; Score 160; DB 5; Length 1738;
 Best Local Similarity 20.9%; Pred. No. 0.016;
 Matches 102; Conservative 88; Mismatches 171; Indels 128; Gaps 21;

QY 54 KDKTAISENL-----KSFSEVGEIDADEVKALTIKIGIKOMKIMMER----- 94
 DB 667 KDNQTINEQLNQLSEKDEKIEKLSNQEQOQDEKINNLLLEIKEDCLIERINQOLLEN 726
 QY 95 -----KEKHTNLMSTLKKCKEKKQKALKLLNEVQEHLEEEER 132
 DB 727 IDLSKYQQLLLEFENFKLNSKEKE-----NQLNELQSKQDERFNQLND--EKLEKEKQ 779
 QY 133 LCRESLADSWGECRSCLENNCRMYTTTCQPSWSSVKNKIERFFKIKYQFLFPFHEDNEKD 192
 DB 780 L--QSIEDFNQYK-----QQQLSSNSN-IDQOQOSTIIESELKEQKELN 822
 QY 193 LPISEKLIIEEDAQLTQMEDVFSQLT-----VDVNSLFNRSFNVFRMQQEQFDTQFSH 245
 DB 823 ---DSKLIEKEKQLQQLQEQFQDLQNEKNQKHQDQLELEKQ---LQALQEQYDQLNETN 876
 QY 246 FISDTDLTEYFFPAFSKEPMTRADLE-----QCWDIPNFOLFQNFVSIVYESV 295
 DB 877 QSIENQNLQNL---INKENLKEQELKLQNLQNOQIEKIQFDQOQEFKQNSINIELV 933
 QY 296 SETITMKLKAIED---LPKQDKA---PDHGLITSKMLPGQDGLCGLEDNLSRCPFHE 349
 DB 934 NERNEKLIQLQDYDQLKQNRNDEKENDLEK--ENQKSIQNELNQLIEKNSDHK 991
 QY 350 KQCKQAHLSDCPD---VPALHTEDE---AIRLVNVSNOQYQIL-----Q 391
 DB 992 EQQLKQOSIENDLIEKENIQQLQSQLNEQROQOSNQLSEKQDQNLQIEKNQFDQKEQ 1051
 QY 392 MTRKHLEDTAY---LVEKMRGFGF-WVSELANQAPETIIFNSIQVVPRIHEGNISKQD 446
 DB 1052 LKQOSIENDLIEKENIQQLQSQLNEQROQOSNQLSEKQDQNLQIEKNESDQK 1105
 QY 447 ETMMTDLSI 455
 DB 1106 EQQLKQOSI 1114

RESULT 10
 Q9W6V0 PRELIMINARY; PRT; 3616 AA.
 ID Q9W6V0
 AC Q9W6V0
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HYPERION PROTEIN, 419 KDA ISOFORM.
 GN HYPERION.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=RETINA;
 RA Kemmerer W.A., Schwarz U.;
 RT "Characterization of Hyperion, a gene coding for an abundance of gene

RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.,
RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT libraries."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB070119; BAB63064.1; -
KW Hypothetical protein.
SQ SEQUENCE 840 AA; 97587 MW; A875627BE3D4D716 CRC64;

Query Match 5.7%; Score 149.5; DB 6; Length 840;
Best Local Similarity 18.6%; Pred. No. 0.034;
Matches 88; Conservative 89; Mismatches 154; Indels 143; Gaps 18;

QY 57 TAIENLKSFEVEGEIDADEVKALGKIKQMKIMMERKEKEHTNLMSTLKKCRE-BKQE 115
DB 426 TEMQKNKVSQY--LEMDKTLKKEEVEKRLQLRKEQKVATASALDLKREKETQOE 483
QY 116 ALKLLNEVQE----HLBEEERLCHESLADSWGECRSCLENNCMRIYTTCPSPWSVKNKI 171
DB 484 FLSQEBFQKRDKANLBERQKL-----KSLR 509
QY 172 ERFRKIYQFLPFPHEDNEKDLPISEKLIEDDAQLTOMEDVFSOLTVDVNSLFNRSF--- 228
DB 510 EKLITQVKNLQFMSENERAKNIKLOQ-----QINVEKNKNKKLQHVARSSEQNTVPK 562
QY 229 NVFRMQOEFQDTQSHFISDTLTPEYFFPAFS---KEPMTKADLE----- 272
DB 563 SETAQLKEOLEEVKMSDITDKTMHNSLLDSCPEESLNPAIDERSQSLAKMSHLL 622
QY 273 -----OCWDIPNFOLFQNFVSIVSVSETITKMLKAI-----EDLPKQDKAP 316
DB 623 ALMVGLLKQDITNSDAE--HFKESS--EKVSDIMQLRLKSLHLKKNLDKELLKHKDRIT 678
QY 317 DHGGLISKMLPGQDRGL---CGE----- 336
DB 679 TFRDLIAKEAFQDHAIKVTDCDSEAKSIRDPVTFGLGAKLDKXVHSLNEEDFLITLKG 738
QY 337 -LDONLSRCFHFHKCQKQAHLSDECDPVPALHTDELDAIRLVNSNQYGGIQLQWTRK 395
DB 739 LSKESHCHNLIEENDKYRHLGLIKKVTSEIEECADQRLAISHSQIAH-LEKRNK 797
QY 396 HLEDTAYLVKMRGQFGWSELANOAPETEIFN---SIQVVPRIHEGNTSKQD 446
DB 798 HLED---LIRKPR-----EKARK-PRSKSLENHKNKMTMPVAFKRNRLD 840

RESULT 13
Q66878 ID 066878 PRELIMINARY; PRT; 1156 AA.
AC 066878
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CHROMOSOME ASSEMBLY PROTEIN HOMOLOG.
GN XCP1 OR A0_629.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujaay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
DR EMBL: AE000699; AAC06839.1; -
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR001687; ATP_GTP_A.
DR InterPro: IPR003716; RNA_pol_Omega.

DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR InterPro: IPR02017; Spectrin.
DR Pfam: PF02483; SMC_C; 1.
DR Pfam: PF02463; SMC_N; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1156 AA; 135563 MW; B12DB30F70C0CE49 CRC64;

Query Match 5.5%; Score 146; DB 16; Length 1156;
Best Local Similarity 20.0%; Pred. No. 0.087;
Matches 89; Conservative 89; Mismatches 162; Indels 104; Gaps 17;

QY 63 LKSEFSEVEGEID----ADEEVKKAITGKIKOMKIMME-----RKEKEHTNLMSTLKK 108
DB 163 IETISGGEYERKEKALELAELVELKIDLEISNOLKRLKEKEKLEKELQ 222
QY 109 CREKQKALKLLNVEQHEERELCRE--SLADSWGECRSCLENNCMRIYTTCPSPWS 166
DB 223 IKRE-TEAKILLKEKELKERERILNELSSRESLEDITFQIQENKEL----- 271
QY 167 VKNKIEFRKIYQFLPFPHEDNEK---DLPISSEKLI-EEDAQLTOMEDVFSOLTVDVNS 222
DB 272 --NERELLKEVNEKIMPFKFKVKTAEIENAEISKEKEREKELSENRYKNLEELINN 329
QY 223 LFNRSFNVRFO---MQOEFQDTQSHFISDTLTPEY-FFPAFSKEPMTKADLEQCWDIP 278
DB 330 LLSKDNLEREVGLTQLELEK-----LKEEYKSLKEVEREKLELEEE----- 373
QY 279 NFFOLFQNF-SVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQDRGLG 337
DB 374 --ERLKITFDEVKLEKEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLE 421
QY 338 DONLSRCFHFHKCQKQAHLSDECDPVPALHTDELDAIRLVNSNQYGGIQLQWTRKHL 397
DB 422 KEDINKLISEREKKEIKKE 481
QY 398 E-----DTAYLVKMRGQFGWSELANOAPETEIFN---SIQVVPRIHEGNTSKQD 423
DB 482 EEVLKEKGAIEREVSFSFSDVDFKDKGVYGSVELIRVKNPEHITAIEVAGGRLKFT 541
QY 424 ----TEIFNSIQVVPRIHEGNTSKQD 443
DB 542 VVEDEEVAKECIQLAKRMNLRGFS 565

RESULT 14
Q28021 ID Q28021 PRELIMINARY; PRT; 1388 AA.
AC Q28021
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RHO-ASSOCIATED KINASE.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OX Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96208507; PubMed=8641286;
RA Matsui T., Amano M., Yamamoto T., Chihara K., Nakafuku M., Ito M.,
RA Nakano T., Okawa K., Iwamatsu A., Kaibuchi K.;
RT "Rho-associated kinase, a novel serine/threonine kinase, as a putative
RT target for small GTP binding protein Rho."
RL EMBO J. 15:2208-2216(1996).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: U36909; AAC48567.1; -
DR HSP: Q63450; IAO6.
DR InterPro: IPR002219; DAG_PE-bind.
DR InterPro: IPR000719; Euk_pkinase.

Query Match 5.5%; Score 144; DB 10; Length 1305;
Best Local Similarity 18.8%; Pred. No. 0.14;
Matches 94; Conservative 82; Mismatches 176; Indels 148; Gaps 18;

Qy	44	LKOSHCAPTWKDKTATSENKLSKSEVEGETADAEVFKALGTGKQIMMERKEKEHTNLM	103
Db	127	LKESHV---KEELFS--LRDIETHQRDSD-----TRASELAQLLESSQQVSDLS	174
Qy	104	STLKCKREEKQEAALKLLNEVQEHLEBEERLCRESLA-----DSWGECRSCLENNCMRY	157
Db	175	ASLKAABEENKAJSSNVETMKNLEQTNTIQELMAELGKLKDSHREKESLSD-----	228
Qy	158	TTQCPSSWSSVKNKIEFPFRKIYQFLPFPHEDNEKDLPISEKLEED-----AOLTM	209
Db	229	-----LVEVETHQRDSDSIHVKELEEQVESKSLVAELNQT	264
Qy	210	-----EDVFSQLTVDVNSLFNRSPNVFRMQOEQDOTFQSHFISDTDLTETPYFPFAPS	262
Db	265	LNNAEBEKKVLSQKIAELSNKEAQNQITQELVSESGQLKESHVKDRDL-----FS	316
Qy	263	-----KEPMTK-ADLEQCHWDIENFQLFCNFVSIVYESVETITMKLKAIEDLPKQ	312
Db	317	LRDIETHQRESSTRVSELEA-----QL-----ESQERISLTDVLDKDAEE	359
Qy	313	DKAPDHGGITSKMLPGQDRGLCGELGDQNLRSRCFPHEKQCKQAHLSDECQDPVPALHTEL	372
Db	360	NKA-----ISSKLEIMDK-----LEQAQNTIKELMDELGELKDRHEK	398
Qy	373	D-EAIRLVNVSNOYQGILOMTKRKHLEDATYLVKMRGQFGVWSELANAQET-----	424
Db	399	ESELSSLVKSADQOQVADMKQSL-----DNAEBEKKMSLO--RILDTISNIEQAQKTIQEH	451
Qy	425	-----EIIFNSQVVPRIHGENISKQDETMMTDLISLPSSNFTLKIPLEES	470
Db	452	MSESELKESHGVKERELTCLRDIETHQRESSTRLESELTQKLLQORVDSLASLNA	511
Qy	471	AESSNFIGYVAKALOHFKE	490
Db	512	EEEKSLSSMILEITDELKO	531

RESULT	15
ID	Q9FJ35
PRELIMINARY;	PRT; 1305 AA.
ID	Q9FJ35
AC	Q9FJ35;
DT	01-MAR-2001 (TtEMBLrel. 16, Created)
DT	01-MAR-2001 (TtEMBLrel. 16, Last sequence update)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: July 2, 2002, 11:44:42 ; Search time 56.51 seconds
(without alignments)
972.952 Million cell updates/sec

Title: US-09-722-544A-2MOD
Perfect score: 2632
Sequence: 1 MKIKAEKNEGSPRSMQLHW.....FIGYVAKALQHFKEHKTKM 495

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 0%

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				A.Geneseq_032802.*			
1:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*	432	16.4	448	12	AA11704	Cytolysis Inhibito
2:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*	405.5	15.4	416	22	AB50285	Apolipoprotein J o
3:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*	400.5	15.2	446	21	AA03441	Porcine clusterin
4:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*	320	12.2	363	22	AAE03764	Human gene 1 encod
5:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*	173.5	6.6	148	22	AA00607	Human polypeptide
6:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*	160.5	6.1	116	21	AA03745	Human secreted pro
7:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*	159.5	6.1	139	22	AAE03783	Human gene 1 encod
8:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*	159.5	6.1	139	22	AAE01703	Human gene 4 encod
9:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*	152.5	5.8	247	22	AAE03787	Human gene 1 encod
10:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*	149	5.7	1372	19	AA56473	Protein with Rho p
11:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*	145	5.5	944	21	AAV67600	Human adipose tiss
12:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*	145	5.5	1388	19	AAW56475	Protein with Rho p
13:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*	143	5.4	1931	22	AB561012	Drosophila melanog
14:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*	141.5	5.4	3433	18	AAW22017	Utrophin. Homo sa
15:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*	140	5.3	934	22	AAU01768	Human secreted pro
16:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*	140	5.3	2517	21	AAV71159	Human phosphodiect
17:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*	139.5	5.3	2688	22	AAW40883	Human polypeptide
18:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*	139	5.3	2633	22	ABG06505	Novel human diago
19:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*	138.5	5.3	2663	22	AAW39097	Human polypeptide
20:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*	138	5.2	5373	22	AAU14603	Human polypeptide
21:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*	138	5.2	5447	22	AAU14697	Novel bone marrow
22:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*	135.5	5.1	1427	12	AAK10534	Human 160kD mediat
23:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2002.DAT.*	135	5.1	1374	22	AAW69070	Human male enhance
24:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2003.DAT.*	133	5.1	990	22	AAW78520	Human protein SEQ
25:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2004.DAT.*	132.5	5.0	1411	17	AAW02258	Nucleolar/endosoma
26:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2005.DAT.*	131	5.0	816	16	AAW66931	AMML chromosome in
27:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2006.DAT.*	130.5	5.0	2482	16	AAW72826	Human mitotin. Ho
28:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2007.DAT.*	130	4.9	1788	22	AAW23996	Human mitotin amin
29:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2008.DAT.*	130	4.9	3248	17	AAW40467	Human polypeptide
30:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2009.DAT.*	130	4.9	885	16	AAW99795	Kinetochores protei
31:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2010.DAT.*	129	4.9	931	22	AAW66930	AMML chromosome in
32:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2011.DAT.*	129	4.9	931	22	AAW79504	Human protein SEQ
33:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2012.DAT.*	128.5	4.9	1392	20	AAW06999	Restin protein seq
34:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2013.DAT.*	127.5	4.8	721	21	AAW21227	Protein encoded by

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2582	98.1	495	20	AAV30785 Protein encoded by
2	2460	93.5	456	20	AAV30793 Immature human HKN
3	2443	92.8	477	20	AAV30786 Protein encoded by
4	2346	89.1	446	20	AAV30792 Mature secreted hum
5	1753.5	66.6	465	20	AAV30791 Bovine HKNG1 ortho
6	1635	62.1	466	20	AAV30787 Protein enoded by
7	1554	59.0	450	20	AAV30788 Guinea pig HKNG1 o
8	1277	48.5	374	20	AAV30789 Guinea pig HKNG1 o
9	1272.5	48.3	373	20	AAV30790 Guinea pig HKNG1 o
10	925.5	35.2	521	20	AAV30794 Amino acid sequenc
11	434	16.5	449	22	AAU28048 Novel human secret

XX New HKNG1 polynucleotides useful in diagnosis and treatment of
PT neuropsychiatric disorders, e.g. bipolar affective disorders and
PT schizophrenia
XX
XX
PS Claim 1; Fig 1A-B; 205pp; English.
XX
CC The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1
CC is a gene associated with bipolar affective disorder (BAD). HKNG1
CC polynucleotides are useful to identify compounds modulating HKNG1 gene
CC expression or HKNG1 polypeptide expression/activity. Compounds inhibiting
CC or enhancing HKNG1 gene expression or activity in individuals can then
CC be administered therapeutically to treat HKNG1-mediated disorders,
CC especially neuropsychiatric disorders e.g. BAD, schizophrenia, or
CC HKNG1-mediated myopia disorders, such as early-onset autosomal
CC dominant myopia. The polynucleotides can be used in gene therapy
CC techniques to treat such disorders. They are also useful in diagnosis
CC to identify individuals having, or at risk of developing, HKNG1-mediated
CC disorders due to mutations in the HKNG1 gene. Such mutations especially
CC result in the production of a protein with a different sequence to
CC the human full-length HKNG1 polypeptide or splice variant sequences,
CC especially the substitution of a lysine for a glutamic acid at residue
CC 202 or 184. The polynucleotides are also useful in gene mapping, to
CC produce probes or primers to identify similar sequences (e.g. mutants
CC or sequences from different species) and to produce transgenic
CC animals.
XX
XX Sequence 495 AA;

Query Match 98.1%; Score 2582; DB 20; Length 495;
Best Local Similarity 99.0%; Pred. No. 1.2e-206;
Matches 489; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MKTKAEKNGPSRWQLHWGDIANNNGNKKPPLLVFIVCLLWLDKSHCAPTWKDKTAITS 60
DB 1 mktaeknegpsrwqlhwgdianngnkkppllvfivclllwkshcaptwkdktaits 60
QY 61 ENLKSFSEGEIDADBEVKALTGIFKIMKIMMERKEHTNLMSTLUKCKREEKQEAALKLL 120
DB 61 enlkfsfsegeidadeevkalktgikmkiimmerkehtnlnstlkkcreekqealkll 120
QY 121 NEVOHLEERELCRSLADSGECRSCLENNCMRIYTTCCPSWSSVKNKIERFERKIYO 180
DB 121 nevgehleeerlcrsladshgcrscleenncmriyttcpswssvknkierfferkiyo 180
QY 181 FLFPFHEDNEKDLPISEKLTIEKDAQLTQMEDVFSQLTVDVNSLNFNSFNVRQMQEFDQ 240
DB 181 flfpfhednekdipiisekltiekdaqltqmedvfsqltvdvnslnfnrsvnrqmqefdq 240
QY 241 TFOSHFISDTDLTEPYFFPAFSKEPTMKADLEOCWDIPNFFQFCNFSVIYESVETIT 300
DB 241 tfghsfisdtdltepyffpafskcptmkadlecwqipnffqfcnfsviyesvsetit 300
QY 301 KMLKATIEDLPKQKADPHGGLISKMLPGODRGICGELDONLSRCFKFHEKQRCQQAHLSE 360
DB 301 kmlkatiedlpkqkadhgglliskmlpgodrgicgeldonslrcfkfhekrcrqqaahlse 360
QY 361 DCPDVPALHTELDEAIRLVNVSNOQYGOILQMTKRKHELDYALVEKMRGQFGWVSELANQ 420
DB 361 dcpdvpalhteldeairlvnvsnoqygoilqmktrkheldtaylvekmrgfgwvselanq 420
QY 421 APETEIIFNISQVPRIHENISKQDETMTDLSILPSSNFTKIPLEESAESSNFIYV 480
DB 421 apeteiifnisqvprihenisksqdetmttdlsilpssnftkipleesaessnfiyv 480
QY 481 VAKALQHFKEHFT 494
DB 481 vakalqhfkehft 494

RESULT 2
AAY30793

ID AAY30793 standard; Protein; 466 AA.
XX
AC AAY30793;
XX
DT 23-NOV-1999 (first entry)
XX
DE Immature human HKNG1 protein form 2.
XX
KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW schizophrenia; splice variant.
OS Homo sapiens.
XX
PN WO9947535-A1.
XX
PD 23-SEP-1999.
XX
PF 16-MAR-1999; 99WO-US05606.
XX
PR 16-MAR-1998; 98US-0078044.
PR 05-JUN-1998; 98US-0088312.
PR 28-OCT-1998; 98US-0106056.
PR 22-JAN-1999; 99US-0236134.
XX
PA (MILL-) MILLENNIUM PHARM INC.
PA (REGC) UNIV CALIFORNIA.
PI Chen H, Freimer NB;
PI
XX
DR WPI; 1999-562047/47.
XX
PT New HKNG1 polynucleotides useful in diagnosis and treatment of
PT neuropsychiatric disorders, e.g. bipolar affective disorders and
PT schizophrenia
XX
PS Claim 22; Fig 17; 205pp; English.

CC The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1
CC is a gene associated with bipolar affective disorder (BAD). HKNG1
CC polynucleotides are useful to identify compounds modulating HKNG1 gene
CC expression or HKNG1 polypeptide expression/activity. Compounds inhibiting
CC or enhancing HKNG1 gene expression or activity in individuals can then
CC be administered therapeutically to treat HKNG1-mediated disorders,
CC especially neuropsychiatric disorders e.g. BAD, schizophrenia, or
CC HKNG1-mediated myopia disorders, such as early-onset autosomal
CC dominant myopia. The polynucleotides can be used in gene therapy
CC techniques to treat such disorders. They are also useful in diagnosis
CC to identify individuals having, or at risk of developing, HKNG1-mediated
CC disorders due to mutations in the HKNG1 gene. Such mutations especially
CC result in the production of a protein with a different sequence to
CC the human full-length HKNG1 polypeptide or splice variant sequences,
CC especially the substitution of a lysine for a glutamic acid at residue
CC 202 or 184. The polynucleotides are also useful in gene mapping, to
CC produce probes or primers to identify similar sequences (e.g. mutants
CC or sequences from different species) and to produce transgenic
CC animals.
XX
SQ Sequence 466 AA;

Query Match 93.5%; Score 2460; DB 20; Length 466;
Best Local Similarity 99.8%; Pred. No. 1.7e-196;
Matches 465; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 30 MKPPLLVFIVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVKKALTGKQMK 89
DB 1 mkppllvfivclllwkshcaptwkdtalsenlksfsevgeldadeevkaltgikqmk 60
QY 90 IMMERKEHTNLMSTLUKCKREEKQEAALKLLNEVQHEERLCLRESLADSGECRSL 149
DB 61 immerkehtnlmstlkkcreekqeaalkllnevqhleeeerlclresladsgecrsl 120

QY 150 ENNCMIYTTCPSSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAOLTM 209
 Db 121 enncmriytcqpswssvknkierfrkikiyqflfpfhednekdlipekliedeaqltm 180
 QY 210 EDVFSQITVDVNSLFRNSFNVRMQOEFDOTFQSHFISDTDLTEPYFFPAFKEPMTKA 269
 Db 181 edvfsqitvdvnslnfrnsfnvrqmqeqdqtfgshfisdtdltepyffpafskpmtka 240
 QY 270 DLEOCWDIPNFOLFCSNFSIYESVSETITKMLKATEDLPKODKAPDHGGLSKMLPGO 329
 Db 241 dleocwdipnflfcsnfsiyesvsetitkmlkatedlpkqakapdhggliskmlpgq 300
 QY 330 DRGLCGELDONLSRCFRFHEKCKQAHLSQEDCPDVPALHTEDEAIRLVNSNOOYGOI 389
 Db 301 drglcgeldnlsrccfrfhekcqahlsqedsedcpvpalhhteldeairlvnsnqyqgi 360
 QY 390 LQWTRKHLEDATYLVKMRGQFGWVSELANOAPETELIFNSIQVVPRIHEGNSKQDET 449
 Db 361 lqmrkhledatylvekmrgqfgwvselanqapetelifnsiqvvprihegniskqdetm 420
 QY 450 MTDLSILPSSNFTLKIPLEESAESSNFIGYVAKALQHKEHFKTW 495
 Db 421 mtdlsilpssnftlkipleesaessnfigyvvakalqhfekhtw 466

RESULT 3
 AAY30786
 ID AAY30786 standard; Protein; 477 AA.
 AC AAY30786;
 XX
 DT 23-NOV-1999 (first entry)
 DE Protein encoded by a human HKNG1 splice variant HKNG1-V1.
 XX HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
 KW schizophrenia; splice variant.
 XX Homo sapiens.
 OS
 XX WO9947535-A1.
 FN
 XX 23-SEP-1999.
 PD
 XX 16-MAR-1999; 99WO-US05606.
 PF
 XX 16-MAR-1998; 98US-0078044.
 PR 05-JUN-1998; 98US-0088312.
 PR 28-OCT-1998; 98US-0106056.
 PR 22-JAN-1999; 99US-0236134.
 XX (MILL-) MILLENNIUM PHARM INC.
 PA (REGC) UNIV CALIFORNIA.
 FI Chen H, Freimer NB;
 XX WPI; 1999-562047/47.
 DR N-PSDB; AAZ10751.
 DR
 XX New HKNG1 polynucleotides useful in diagnosis and treatment of
 PT neuropsychiatric disorders, e.g. bipolar affective disorders and
 PT schizophrenia -
 XX Claim 1; Fig 2A-B; 205pp; English.
 PS
 XX The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)
 CC splice variant. HKNG1 is a gene associated with bipolar affective
 CC disorder (BAD). HKNG1 polynucleotides are useful to identify compounds
 CC modulating HKNG1 gene expression or HKNG1 polypeptide expression/
 CC activity. Compounds inhibiting or enhancing HKNG1 gene expression or
 CC activity in individuals can then be administered therapeutically to
 CC treat HKNG1-mediated disorders, especially neuropsychiatric disorders

CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as
 CC early-onset autosomal dominant myopia. The polynucleotides can be used
 CC in gene therapy techniques to treat such disorders. They are also useful
 CC in diagnosis to identify individuals having, or at risk of developing,
 CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such
 CC mutations especially result in the production of a protein with a
 CC different sequence to the human full-length HKNG1 polypeptide or
 CC splice variant sequences, especially the substitution of a lysine for
 CC a glutamic acid at residue 202 or 184. The polynucleotides are also
 CC useful in gene mapping, to produce probes or primers to identify
 CC similar sequences (e.g. mutants or sequences from different species)
 CC and to produce transgenic animals.
 XX
 SQ Sequence 477 AA;

Query Match 92.8%; Score 2443; DB 20; Length 477;
 Best Local Similarity 98.9%; Pred. No. 4.5e-195;
 Matches 465; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 25 NNSGNMKPPLLVTVCLLWLLKDSHCAPTWKDKTAISENLKSFSEVGIDEVVKALTG 84
 Db 7 nsngnmkppllvtfvclllwkdschaptwkdtaisenlksfsevgideadeevkaltg 66
 QY 85 IKQKIMMERKEKEHTNLMSTLKKCREKQKALKLLNEVQEHLEEEERLCRESLADSWGE 144
 Db 67 ikqkimmerkekehtnlmstlkkcreekqalkllnevqehleeeerlcresladswge 126
 QY 145 CRSCLENNCMRIYTTQPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDA 204
 Db 127 crsclenncmriyttcqpsswssvknkierfrkiyqflfpfhednekdlipeklieda 186
 QY 205 QLQTMEDVFSQLTVDVNSLFRNSFNVRMQOEFDOTFQSHFISDTDLTEPYFFPAFSKE 264
 Db 187 qlqtmdevfsqldvvnslfrnsfnvrqmqeqdqtfgskfisdtdltepyffpafske 246
 QY 265 PMTKADLEOCWDIPNFQFQFNFVSIVSETITKMLKAIEDLPKQDKAPDHGGLISK 324
 Db 247 pmtkadlegcwidpwwffqlfcnfsvsiyesvsetitkmlkaiedlpkqdkapdhggli 306
 QY 325 MLPGQDRGLCGELDONLSRCFRFHEKCKQAHLSQEDCPDVPALHTEDEAIRLVNSNQ 384
 Db 307 mlpgqdrglcgeldnlsrccfrfhekcqahlsqedsedcpvpalhhteldeairlvnsnq 366
 QY 385 QYGQILQWTRKHLEDATYLVKMRGQFGWVSELANOAPETELIFNSIQVVPRIHEGNSK 444
 Db 367 qygqilqmrkhledatylvekmrgqfgwvselanqapetelifnsiqvvprihegnisk 426
 QY 445 QDETMMTDLSILPSSNFTLKIPLEESAESSNFIGYVAKALQHKEHFKT 494
 Db 427 qdetmtdlsilpssnftlkipleesaessnfigyvvakalqhfekht 476

RESULT 4
 AAY30792
 ID AAY30792 standard; Protein; 446 AA.
 AC AAY30792;
 XX
 DT 23-NOV-1999 (first entry)
 DE Mature secreted human HKNG1 protein sequence.
 XX
 KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
 KW schizophrenia; splice variant.
 XX Homo sapiens.
 OS
 XX WO9947535-A1.
 PN
 XX 23-SEP-1999.
 PD
 XX

PF	16-MAR-1999;	99WO-US05606.	
XX	16-MAR-1998;	98US-0078044.	
PR	05-JUN-1998;	98US-0088312.	
PR	28-OCT-1998;	98US-0106056.	
PR	22-JAN-1999;	99US-0236134.	
XX	(MILL-) MILLENNIUM PHARM INC.		
PA	(REGC) UNIV CALIFORNIA.		
XX	Chen H, Frelmer NB;		
XX	WPI; 1999-562047/47.		
XX	New HKNG1 polynucleotides useful in diagnosis and treatment of		
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and		
PT	schizophrenia		
XX	Claim 5; Fig 17; 205pp; English.		
XX	The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1		
CC	is a gene associated with bipolar affective disorder (BAD). HKNG1		
CC	polynucleotides are useful to identify compounds modulating HKNG1 gene		
CC	expression or HKNG1 polypeptide expression/activity. Compounds inhibiting		
CC	or enhancing HKNG1 gene expression or activity in individuals can then		
CC	be administered therapeutically to treat HKNG1-mediated disorders,		
CC	especially neuropsychiatric disorders e.g. BAD, schizophrenia, or		
CC	HKNG1-mediated myopia disorders, such as early-onset autosomal		
CC	dominant myopia. The polynucleotides can be used in gene therapy		
CC	techniques to treat such disorders. They are also useful in diagnosis		
CC	to identify individuals having, or at risk of developing, HKNG1-mediated		
CC	disorders due to mutations in the HKNG1 gene. Such mutations especially		
CC	result in the production of a protein with a different sequence to		
CC	the human full-length HKNG1 polypeptide or splice variant sequences,		
CC	especially the substitution of a lysine for a glutamic acid at residue		
CC	202 or 184. The polynucleotides are also useful in gene mapping, to		
CC	produce probes or primers to identify similar sequences (e.g. mutants		
CC	or sequences from different species) and to produce transgenic		
CC	animals.		
XX	Sequence 446 AA;		
SQ			
Query Match 89.1%; Score 2346; DB 20; Length 446;			
Best Local Similarity 99.8%; Pred. No. 5e-187; Indels 0; Gaps 0;			
Matches 445; Conservative 1; Mismatches 0;			
QY	50	APTWKDTAISENLKSFSEVGEIDADEEVKALTGKIMMERKEKEHTLMSTLKKC	109
DB	1	aptwkdtaisenklsfsevgeladeevkaltgikqmkimmerkekehtlmstlkkc	60
QY	110	REEKQALKLLNEVQHLEERLCHRESLADSWGECRSCLNENCMRIYTCQPSWSSVKN	169
DB	61	reekqalkllnevqhleerlchresladswwgecrsclenncmriyttcqpssvsn	120
QY	170	KIERFRKIYQIFLPPHEDNEKDLPISEKLIKDAQLTQMEDVFSQTLVDVNSLNRSPN	229
DB	121	kierfrkiyqiflphednekdipiseklicedaqltqmedvfsqtlvdvnslnrfsn	180
QY	230	VFRMQQEDTQTFQSHFISDITLTPYFPFAPFSKEPMTKADLEQCDWIPNFQFCNFESV	289
DB	181	virmqqeqdqtqsfhfsdtdltpeyffpafskpmtkadleqcdwipnfqfcnfsv	240
QY	290	SIYESVSEITKMLKAIEDLPQDKAPDHGGLTSKMLPCQDRGLGCELQDNLSRCFKPHE	349
DB	241	siyevseitkmlkaiedlpqdkapdhggllskmlpgqdrglgcelqdnlsrckfthe	300
QY	350	KCKQCOAHLSEDCPDVPAHLTELDEAIRLVNVSNOQYGOILQMTKRKHLDETAYLVEKMRG	409
DB	301	kcqkcahlsecdpvpahlteldeairlvnvsnoqyqgilqmtkrkhldeatylvekmrg	360
QY	410	QFGWVSELANQAPETEIFNSIQVPRTHEGNISQDETMTDLSILPSSNFTLKIPLEE	469

Db	361	qfgwvselanqapetelifnsiqvprthehgnlskqdetmtdlsilpsnftlkiplee	420
QY	470	SAESSNFIGYVWAKALQHKEHFKTW	495
DB	421	saessnfigyvvakalqhkehfkwtw	446
RESULT	5		
AAV30791			
ID	AAV30791	standard; Protein; 465 AA.	
XX	AAV30791;		
AC	AAV30791;		
XX	23-NOV-1999	(first entry)	
DT	23-NOV-1999	(first entry)	
XX	Bovine HKNG1 ortholog splice variant protein.		
DE	Bovine HKNG1 ortholog splice variant protein.		
XX	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;		
KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;		
KW	schizophrenia; splice variant.		
XX	Bos sp.		
XX	WO9947535-A1.		
PN	23-SEP-1999.		
XX	16-MAR-1999;	99WO-US05606.	
XX	16-MAR-1998;	98US-0078044.	
PR	05-JUN-1998;	98US-0088312.	
PR	28-OCT-1998;	98US-0106056.	
PR	22-JAN-1999;	99US-0236134.	
XX	(MILL-) MILLENNIUM PHARM INC.		
PA	(REGC) UNIV CALIFORNIA.		
XX	Chen H, Frelmer NB;		
PI	WPI; 1999-562047/47.		
DR	N-PSDB; AAZ10759, AAZ10760, AAZ10761.		
XX	New HKNG1 polynucleotides useful in diagnosis and treatment of		
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and		
PT	schizophrenia		
XX	Claim 1; Fig 11A-B; 205pp; English.		
XX	The present sequence is encoded by bovine HKNG1 ortholog splice		
CC	variant. HKNG1 (Hong Kong new gene 1) is a gene associated		
CC	with bipolar affective disorder (BAD). HKNG1 polynucleotides are		
CC	useful to identify compounds modulating HKNG1 gene expression or		
CC	HKNG1 polypeptide expression/activity. Compounds inhibiting or		
CC	enhancing HKNG1 gene expression or activity in individuals can		
CC	then be administered therapeutically to treat HKNG1-mediated		
CC	disorders, especially neuropsychiatric disorders e.g. BAD,		
CC	schizophrenia, or HKNG1-mediated myopia disorders, such as		
CC	early-onset autosomal dominant myopia. The polynucleotides can be used		
CC	in gene therapy techniques to treat such disorders. They are also useful		
CC	in diagnosis to identify individuals having, or at risk of developing,		
CC	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such		
CC	mutations especially result in the production of a protein with a		
CC	different sequence to the human full-length HKNG1 polypeptide or		
CC	splice variant sequences, especially the substitution of a lysine for		
CC	a glutamic acid at residue 202 or 184. The polynucleotides are also		
CC	useful in gene mapping, to produce probes or primers to identify		
CC	similar sequences (e.g. mutants or sequences from different species)		
CC	and to produce transgenic animals.		
XX	Sequence 465 AA;		
SQ			
Query Match 66.6%; Score 1753.5; DB 20; Length 465;			

Best Local Similarity 70.8%; Pred. No. 1.3e-137;
Matches 330; Conservative 58; Mismatches 77; Indels 1; Gaps 1;

```
QY 30 MKPPLVIVCLLWKDSHCAPTWKDKTAISENLKSFSEVGEIDAEVKKALGIQOMK 89
D 1 mkppllvivllqrdccaptgkdrtsredpkgsageldvdeevkallgmkgm 60
QY 90 IMMERKEKEHTNLMTLKKREEKQKALKLLNEVQHELEERLCRESLADSWGECRSL 149
D 61 ilmerreehsklmrtlkckreekqalklmevqehleeeerlcqylngswdeckscl 120
QY 150 ENNCMRIYTCOPSWSSVKNKIERFRKIYQFLPFHEDNEKDLPISEKLIKEDAQLTQM 209
D 121 esdcmrfyttcqswwsmkstiervfrkiyqlfipfheddekelpvgekfteedvqlmqi 180
QY 210 EDVFSQLTVDVNSLFRNSFNVRQMQEFDQTFQSHFISDLDLTPYFFPAFSKEPMTKA 269
D 181 envfsqtlvdvglynnshfvkqmqdelfafqsfmsdtdsmepffafskpaka 240
QY 270 DLEQCWDIPNPFQFCNFSVSIYESVETITKMLKAIEDLPKQKAPDHGGLISKMLPGQ 329
D 241 hpmqswdipstfifcnfslsvyqsvsatvtemikaiedlskqdsahgppsattvpr 300
QY 330 DRGLCGELDQNLSCFCFHEKCKQCOAHLSEDCPDVPAALHTELDEATRLVNVNSQQYGOI 389
D 301 grgicgepgqnsseclgharcqkcdylwadcpavpelytkadealelvaisnqgyavq 360
QY 390 LQMTRKHLEDYALVLEKMRGQFGWVSELANOAPETEIFNSIQVVPRI--HEGNSIKQDET 449
D 361 lqmtqhledtymeknregfvgwtelascpgnsenifsfikvvpvgvhegnfskqdek 420
QY 450 MTDLSILPSSNFTLIKPILEESAESSNFIYGYVAKALQHFKEHFTW 495
D 421 i-disilpssnftitipleesaessdfisymalakavghfkehfksw 465
```

RESULT 6

AA30787
ID AAY30787 standard; Protein; 466 AA.

AC AAY30787;

XX AAY30787;

DT 23-NOV-1999 (first entry)

XX Protein emoded by the guinea pig HKNG1 ortholog gphkng1815.

DE HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW schizophrenia; splice variant.

OS Cavia cobaya.

XX WO9947535-A1.

PN 23-SEP-1999.

XX 16-MAR-1999; 99WO-US05606.

PF 16-MAR-1998; 98US-0078044.

PR 05-JUN-1998; 98US-0088312.

PR 28-OCT-1998; 98US-0106056.

PR 22-JAN-1999; 99US-0236134.

XX (MILL-) MILLENNIUM PHARM INC.
PA (REGC) UNIV CALIFORNIA.

XX Chen H, Freimer NB;

XX WPT; 1999-562047/47.

DR N-PSDB; AA210755.

XX New HKNG1 polynucleotides useful in diagnosis and treatment of
PT neuropsychiatric disorders, e.g. bipolar affective disorders and

PT schizophrenia -
XX Claim 1; Fig 7A-B; 205pp; English.

XX The present sequence is encoded by guinea pig HKNG1 ortholog gphkng1815.
CC HKNG1 (Hong Kong new gene 1) is a gene associated with bipolar affective
CC disorder (BAD). HKNG1 polynucleotides are useful to identify compounds
CC modulating HKNG1 gene expression or HKNG1 polypeptide expression/
CC activity. Compounds inhibiting or enhancing HKNG1 gene expression or
CC activity in individuals can then be administered therapeutically to
CC treat HKNG1-mediated disorders, especially neuropsychiatric disorders
CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia. The polynucleotides can be used
CC in gene therapy techniques to treat such disorders. They are also useful
CC in diagnosis to identify individuals having, or at risk of developing,
CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such
CC mutations especially result in the production of a protein with a
CC different sequence to the human full-length HKNG1 polypeptide or
CC splice variant sequences, especially the substitution of a lysine for
CC a glutamic acid at residue 202 or 184. The polynucleotides are also
CC useful in gene mapping, to produce probes or primers to identify
CC similar sequences (e.g. mutants or sequences from different species)
CC and to produce transgenic animals.

XX Sequence 466 AA;

Query Match 62.1%; Score 1635; DB 20; Length 466;
Best Local Similarity 66.5%; Pred. No. 9.5e-128;
Matches 311; Conservative 64; Mismatches 89; Indels 4; Gaps 3;

```
QY 30 MKPPLVIVCLLWKDSHCAPTWKDKTAISENLKSFSEVGEIDAEVKKALGIQOMK 89
D 1 mkpllmfpvcllwkdchcaptwkdktaisenansfseageidvdeevkallgikgm 60
QY 90 IMMERKEKEHTNLMTLKKREEKQKALKLLNEVQHELEERLCRESLADSWGECRSL 149
D 61 immerreehsklmrtlkckreekqalklmevqehleeeerlcqylngswdeckscl 120
QY 150 ENNCMRIYTCOPSWSSVKNKIERFRKIYQFLPFHEDNEKDLPISEKLIKEDAQLTQM 209
D 121 esncmrfdtctcpawssvknmveqfrrkiyqlfipqge-ndrsrgpvskgvtdeeqvshi 179
QY 210 EDVFSQLTVDVNSLFRNSFNVRQMQEFDQTFQSHFISDLDLTPYFFPAFSKEPMTKA 269
D 180 ehvfsqsladvtslfnrslvyfklrrfcdafsfyftsgtdvtepfifalskepaya 239
QY 270 DLEQCWDIPNPFQFCNFSVSIYESVETITKMLKAIEDLPKQKAPDHGGLISKMLPGQ 329
D 240 daepswaiipnvfqlcnlsfsvyqsvsekltitratedppkqkdsngggsikilpeq 299
QY 330 DRGLCGELDQNLSCFCFHEKCKQCOAHLSEDCPDVPAALHTELDEATRLVNVNSQQYGOI 389
D 300 drgsdglgqnlscdvnfkrccqkcdyisddcpnvpeylrelnealrfsrsnqqyqv 359
QY 390 LQMTRKHLEDYALVLEKMRGQFGWVSELANOAPETEIFNSIQVVPRI--HEGNSIKODE 447
D 360 vqmtqyhledtllmekmregfvgwvselayqspgaedifnpykvmvalsahegnssddq 419
QY 448 TMTDLSILPSSNFTLIKPILEESAESSNFIYGYVAKALQHFKEHFTW 495
D 420 tvvps-silpssnftlaspleksagnanfidhvvkvqlghfkehfkfw 466
```

RESULT 7

AA30788
ID AAY30788 standard; Protein; 450 AA.

XX AAY30788;

XX 23-NOV-1999 (first entry)

XX Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7b.

CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such
CC mutations especially result in the production of a protein with a
CC different sequence to the human full-length HKNG1 polypeptide or
CC splice variant sequences, especially the substitution of a lysine for
CC a glutamic acid at residue 202 or 184. The polynucleotides are also
CC useful in gene mapping, to produce probes or primers to identify
CC similar sequences (e.g. mutants or sequences from different species)
XX and to produce transgenic animals.
XX
SQ Sequence 374 AA;

Query Match 48.5%; Score 1277; DB 20; Length 374;
Best Local Similarity 54.5%; Pred. No. 4.8e-98;
Matches 255; Conservative 44; Mismatches 73; Indels 96; Gaps 3;

Qy 30 MKPPLLVIVCLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDADEVKKALGTGKQMK 89
Db 1 mkpllpvclllwkchcaptwkdktaisenansfseageldvdevkialigikmqk 60
Qy 90 IMMERKEHTNLMSTLKKREEKQKALKLLNEVQHEERLCRESLADSWGECRSCL 149
Db 61 immerreehsklmktlkkckeeqalklmnevheheeslcqvsldswdecrac 120
Qy 150 ENNCMRIYTCOPSSWSSVKNKIERFRKIYQFLPFFHEDNEKDLPISEKLEKDAQLTQM 209
Db 121 esncmrfdttcqpawssvkn----- 140
Qy 210 EDVFSQLTVDVNSLNFNSFNVRQMQQEFDTQFSHFISDITDLETPYFFPAFSKEPTKA 269
Db 141 -----mepayra 147
Qy 270 DLEQCWDIPNFQFCNFSVSIYESVETITKMLKAIEDLPKQDKAPDHGGLSKMLPGQ 329
Db 148 daepswaipnvqllcnlsfsvysvseklittratedppkqkdsngqgppiskilpeq 207
Qy 330 DRGLCGELDQNLSCRFKHEKQKQCAHLSEDCPDVPALHTELDEAIRLVNVSNOQYQGI 389
Db 208 drgsdgklqnlscvnrfrckqckqylsddcpnvpeylrelnealrlvsrsnqyqdv 267
Qy 390 LQMTKRLHEDTAYLVEKMRGQFQWVSELANOAPETEIFNSIOVVPRI--HEGNISKODE 447
Db 268 vqmtqyhlledtllimekmrefgwwselaydspgaedlnfpvkvmlvalsahegnssqdd 327
Qy 448 TMTDLSLTPSNFTLKLIPLESSESSNFYGYVYAKALQHFKEFTKW 495
Db 328 tvvps-sllpssntflspleksagnanfidhvvekvqlhfkelnktw 374

RESULT 9
AAY30790
ID AAY30790 standard; Protein; 373 AA.
XX
AC AAY30790;
XX
DT 23-NOV-1999 (first entry)
XX
DE Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7d.
XX
KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW schizophrenia; splice variant.
XX
OS Cavia cobaya.
XX
PN W09947535-A1.
XX
PD 23-SEP-1999.
XX
PF 16-MAR-1999; 99WO-US05606.
XX
PR 16-MAR-1998; 98US-0078044.
PR 05-JUN-1998; 98US-0088312.

PR 28-OCT-1998; 98US-0106056.
PR 22-JAN-1999; 99US-0236134.
XX
PA (MILL-) MILLENNIUM PHARM INC.
PA (REGC) UNIV CALIFORNIA.
XX
PI Chen H, Freimer NB;
XX
DR WPI; 1999-562047/47.
DR N-PSDB; AA210758.
XX
PT New HKNG1 polynucleotides useful in diagnosis and treatment of
PT neuropsychiatric disorders, e.g. bipolar affective disorders and
PT schizophrenia -
XX
PS Claim 1; Fig 10A-B; 205pp; English.
XX
CC The present sequence is encoded by a guinea pig HKNG1 ortholog
CC gphkng1815 splice variant gphkng7d. HKNG1 (Hong Kong new gene 1)
CC is a gene associated with bipolar affective disorder (BAD). HKNG1
CC polynucleotides are useful to identify compounds modulating HKNG1
CC gene expression or HKNG1 polypeptide expression/activity. Compounds
CC inhibiting or enhancing HKNG1 gene expression or activity in
CC individuals can then be administered therapeutically to treat
CC HKNG1-mediated disorders, especially neuropsychiatric disorders
CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as
CC early-onset autosomal dominant myopia. The polynucleotides can be used
CC in gene therapy techniques to treat such disorders. They are also useful
CC in diagnosis to identify individuals having, or at risk of developing,
CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such
CC mutations especially result in the production of a protein with a
CC different sequence to the human full-length HKNG1 polypeptide or
CC splice variant sequences, especially the substitution of a lysine for
CC a glutamic acid at residue 202 or 184. The polynucleotides are also
CC useful in gene mapping, to produce probes or primers to identify
CC similar sequences (e.g. mutants or sequences from different species)
CC and to produce transgenic animals.
XX
SQ Sequence 373 AA;

Query Match 48.3%; Score 1272.5; DB 20; Length 373;
Best Local Similarity 54.5%; Pred. No. 1.1e-97;
Matches 255; Conservative 45; Mismatches 71; Indels 97; Gaps 4;

Qy 30 MKPPLLVIVCLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDADEVKKALGTGKQMK 89
Db 1 mkpllpvclllwkchcaptwkdktaisenansfseageldvdevkialigikmqk 60
Qy 90 IMMERKEHTNLMSTLKKREEKQKALKLLNEVQHEERLCRESLADSWGECRSCL 149
Db 61 immerreehsklmktlkkckeeqalklmnevheheeslcqvsldswdecrac 120
Qy 150 ENNCMRIYTCOPSSWSSVKNKIERFRKIYQFLPFFHEDNEKDLPISEKLEKDAQLTQM 209
Db 121 esncmrfdttcqpawssvkn----- 140
Qy 210 EDVFSQLTVDVNSLNFNSFNVRQMQQEFDTQFSHFISDITDLETPYFFPAFSKEPTKA 269
Db 141 -----mepayra 146
Qy 270 DLEQCWDIPNFQFCNFSVSIYESVETITKMLKAIEDLPKQDKAPDHGGLSKMLPGQ 329
Db 147 daepswaipnvqllcnlsfsvysvseklittratedppkqkdsngqgppiskilpeq 206
Qy 330 DRGLCGELDQNLSCRFKHEKQKQCAHLSEDCPDVPALHTELDEAIRLVNVSNOQYQGI 389
Db 207 drgsdgklqnlscvnrfrckqckqylsddcpnvpeylrelnealrlvsrsnqyqdv 266
Qy 390 LQMTKRLHEDTAYLVEKMRGQFQWVSELANOAPETEIFNSIOVVPRI--HEGNISKODE 447
Db 267 vqmtqyhlledtllimekmrefgwwselaydspgaedlnfpvkvmlvalsahegnssqdd 326

QY	448	TMWTDLSILPSSNFTLKIPLESAESSNFTGYVYVAKALQHFKEKWTW	495
Db	327	tvtps-sllpssnftlsspleksagnanfidhvvkvlqhfkefkwc	373
RESULT	10		
AA30794			
ID	AA30794	standard; Protein; 521 AA.	
XX	AA30794;		
XX	23-NOV-1999	(first entry)	
DE	Amino acid sequence of human HKNG1 splice variant HKNG1-delta7.		
XX	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;		
KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;		
KW	schizophrenia; splice variant.		
XX	Homo sapiens.		
OS	WO9947535-A1.		
PN	23-SEP-1999.		
PD	16-MAR-1999;	99WO-US05606.	
XX	16-MAR-1998;	98US-0078044.	
PR	05-JUN-1998;	98US-0088312.	
PR	28-OCT-1998;	98US-0106056.	
PR	22-JAN-1999;	99US-0236134.	
XX	(MILL-) MILLENNIUM PHARM INC.		
PA	(REGC) UNIV CALIFORNIA.		
XX	Chen H, Freimer NB;		
PI	WPI; 1999-562047/47.		
DR	N-PSDB; AA210762.		
XX	New HKNG1 polynucleotides useful in diagnosis and treatment of		
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and		
PT	schizophrenia		
XX	Claim 1; Fig 18A-B; 205pp; English.		
PS	The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)		
XX	splice variant. HKNG1 is a gene associated with bipolar affective		
CC	disorder (BAD). HKNG1 polynucleotides are useful to identify compounds		
CC	modulating HKNG1 gene expression or HKNG1 polypeptide expression/		
CC	activity. Compounds inhibiting or enhancing HKNG1 gene expression or		
CC	activity in individuals can then be administered therapeutically to		
CC	treat HKNG1-mediated disorders, especially neuropsychiatric disorders		
CC	e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as		
CC	early-onset autosomal dominant myopia. The polynucleotides can be used		
CC	in gene therapy techniques to treat such disorders. They are also useful		
CC	in diagnosis to identify individuals having, or at risk of developing,		
CC	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such		
CC	mutations especially result in the production of a protein with a		
CC	different sequence to the human full-length HKNG1 polypeptide or		
CC	splice variant sequences, especially the substitution of a lysine for		
CC	a glutamic acid at residue 202 or 184. The polynucleotides are also		
CC	useful in gene mapping, to produce probes or primers to identify		
CC	similar sequences (e.g. mutants or sequences from different species)		
CC	and to produce transgenic animals.		
XX	Sequence	521 AA;	
SQ	Query Match	35.2%;	Score 925.5; DB 20;
	Best Local Similarity	77.0%;	Pred. No. 1.5e-68;
	Matches 181; Conservative	7;	Mismatches 16;
		Indels	31; Gaps 3;

QY	1	MKIAKNEGSPSRWQLHWGDIANNNGNKKPDLVFIIVCLLWLDKSHCAPTWKDKTAIS	60
Db	94	mkiaeknegpsrswqlhwgdianngnmkppllvfvicllwldshcaphwkdktais	153
QY	61	ENLKSFEVGEIDAEVKKALTGKQKIMMERKEKEHTNLMSTLKKCREEKQKALKLL	120
Db	154	enlksfsevgeldadeevkkaltgikqkimmerkekehcnlmstlkkcreekqkall	213
QY	121	NEVOEHLEEEERLCRESLADSWGECRSCLENMCMRYTTTCOPSSWSSVKNKI--ERFFRKI	178
Db	214	nevgehleeerleeresladsweocrscleenncmrlytttcqpswssvknklltteafqrc	273
QY	179	YQFLFFPHEDNEKDLPISEKLEIEKDAQLTQMEDVFSOLT---VDVNSLFFNRSNV	230
Db	274	y-----lgrtedcvgnltricqdvsnfmknknv	302
RESULT	11		
AA28048			
ID	AAU28048	standard; Protein; 449 AA.	
XX	AAU28048;		
XX	18-DEC-2001	(first entry)	
DE	Novel human secretory protein, Seq ID No 217.		
XX	Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;		
KW	ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;		
KW	transgenic animal; Alzheimer's disease; Parkinson's disease; burn;		
KW	amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;		
KW	ulcer; osteoporosis; bone degenerative disorder; periodontal disease;		
KW	gut protection; lung; liver fibrosis; immune deficiency; infection;		
KW	severe combined immunodeficiency; SCID; autoimmune disorder; allergy;		
KW	multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;		
KW	fertility; analgesic; pain; antigen.		
XX	Homo sapiens.		
OS	WO200166689-A2.		
PN	13-SEP-2001.		
XX	05-MAR-2001;	2001WO-US04942.	
XX	07-MAR-2000;	2000US-0519705.	
PR	19-MAY-2000;	2000US-0574454.	
PR	17-JUN-2000;	2000US-0596193.	
PR	14-JUL-2000;	2000US-0616847.	
PR	19-SEP-2000;	2000US-0653263.	
PR	20-OCT-2000;	2000US-0693267.	
XX	(HYSE-) HYSEQ INC.		
PA	Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;		
PI	Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;		
XX	WPI; 2001-589334/66.		
DR	N-PSDB; AA544948.		
XX	Novel polypeptides and nucleic acids obtained from cDNA libraries		
PT	prepared from various human tissues, for diagnosis and treatment of		
PT	cancer, neurological, inflammatory, and autoimmune disorders -		
XX	Example 3; SEQ ID No 217; 107pp; English.		
PS	The invention relates to novel isolated human secreted polypeptides (I)		
CC	and polynucleotides (II). (I) and (II) are useful for treating		
CC	inflammatory conditions such as arthritis, nephritis, Crohn's disease,		
CC	ischaemia-reperfusion injury, shock, sepsis, immune responses, and is		
CC	involved in increasing haematopoiesis, stem cell survival, bone growth		
CC	and remodeling. (I), (II) and modulators of (II) are useful for		
CC	prophylaxis or treatment of one or more cancers. (II) is also useful for		

CC creating transgenic animals useful for studying the in vivo activities of
CC the polypeptide as well as for studying modulators of the polypeptides.
CC (1) induces the proliferation of neural cells and regeneration of nerve
CC and brain tissue and is useful for the treatment of central and
CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,
CC Parkinson's disease, Huntington's disease, and amyotrophic lateral
CC sclerosis. In addition, (1) is involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis and is useful for treating myeloid
CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC tissue growth, and in tissue repair, healing of burns, incisions,
CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative
CC disorders, or periodontal disease. Furthermore, (1) is also useful for
CC gut protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, various immune deficiencies and
CC disorders including severe combined immunodeficiency (SCID), bacterial or
CC fungal infections, autoimmune disorders e.g. multiple sclerosis,
CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic
CC reactions and conditions, such as asthma or other respiratory problems.
CC In addition, (1) affects biorhythms or circadian cycles of rhythms,
CC fertility, metabolism, catabolism, anabolism, storage or elimination of
CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides
CC analgesic effects or other pain reducing effects, immunoglobulin like
CC activity and can act as an antigen in a vaccine composition to raise an
CC immune response. AAU28020-AAU28395 represent novel human secreted protein
CC amino acid sequences of the invention.
XX
SQ Sequence 448 AA;

Query Match 16.5%; Score 434; DB 22; Length 449;
Best Local Similarity 25.8%; Pred. No. 1.1e-27;
Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

QY 30 MKPPLLYFIVCLLWLDKSHCAPTWK-----DKTAISENLKSFSEVGEIDAEVKKALT 83
Db 1 mmtkllilfvgl-----twesqvgldgtvsnelqemsnqgskynkeiqnavn 51
QY 84 GIKQMKIMMERKEHTNLMSTLKKCKREEKQALKLLNEVOHEERLCRESLADSWG 143
Db 52 gvkqiktneerktllnleaaakkkedalnretsetkkelpgvcnetmmalwe 111
QY 144 ECRSCLENNCRIVT-TCQPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIK 202
Db 112 eckpcklqctcmkfyarvcrgsglvgqlqleeflnqspfyfwmgdr-----idsllen 165
QY 203 DAQLTQMEDV----FSQLTVDVNSLNFNRSNVFMQMQEFDQTFQSHFIS-DTDLTEPYF 257
Db 166 drqqtldvmqdhfrassildefqdrf-----ftrepqdy--hylpfsphrrphf 218
QY 258 PPAFSKPEMTKADLE-QCWDIPNPFQFCNFSVSIYESVSFTITKMLKATEDLPKQKAP 316
Db 219 f--fpksrivrslmpfpyepelnhamfpflemihe-----qqamdlfhspaf 267
QY 317 DHGGLSKMLPGQDGLGELDQNLSCRFKPFHEKQRCQAHSEDC----PDVPALHTE 372
Db 268 qhpptefiregddrtvcirehnstgclrmkdcqckreillsvdcstnnpqaklrel 327
QY 373 DEATRLVNSVNOQYQIILQWTRKHELDATYLVKMRQFGVHSELANOAPETEIFNSIQ 432
Db 328 deslqvaerlrlkynellskysqkmlntssilleqlneqfnwvrsrlantqgdqdyrrvt 387
QY 433 VVPRITHEGNISKQDETMTDLSILPSSNFTLKIPLIESAESNSFVGVAKALQHF-KEH 491
Db 388 tvas-htsdsdvpgsvtevvvklfidsdpitvtpvpsvrknkpfmetvaealqeyrkkh 446

RESULT 12

AAR11704

ID AAR11704 standard; Protein; 448 AA.

XX

AC AAR11704;

XX

DT 20-JUN-1991 (first entry)
XX Cytolysis Inhibitor.
DE Cytolysis Inhibitor.
XX
KW cytotoxic inhibitor; perforin; immunological effector molecule;
KW infertility.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..21 /label= signal peptide
FT Protein 22..226 /label= A-chain
FT Protein 227..448 /label= B-chain
XX DE3933850-A.
PN 18-APR-1991.
XX
PD 06-OCT-1989; 89DE-3933850.
XX
PR 06-OCT-1989; 89DE-3933850.
XX (SCHD) SCHERING AG.
XX
PI Tschopp J, Jenne D;
XX
DR WPI; 1991-118338/17.
XX
PT DNA sequence coding for cytotoxic inhibitor - is strong inhibitor
PT of terminal complement protein, eg perforin secreted by killer
PT cells
XX
PS Claim 13; Page 9; 15pp; German.
XX
CC This cytotoxic inhibitor is encoded by a 1.7kb BamHI-KpnI fragment
CC isolated from a liver-specific cDNA library. It is a blood plasma
CC component that inhibits immunological effector molecules. It is used
CC for systemic or local treatment of inflammatory or autoimmune
CC diseases mediated by complement or killer cells. It can also be
CC used for detoxification of membrane-active and cytolytic proteins
CC released by bacteria, fungi and insect venoms. The protein has a
CC further use in the treatment of infertility caused by deficiency of
CC cytotoxic inhibitor. Monoclonal antibodies directed against the
CC cytotoxic inhibitor are also covered by the invention. They are used
CC to isolate or quantify the natural protein in human plasma.
CC See also AAQ11501 and AAQ11502.
XX
SQ Sequence 448 AA;

Query Match 16.4%; Score 432; DB 12; Length 448;
Best Local Similarity 25.8%; Pred. No. 1.1e-27;
Matches 123; Conservative 96; Mismatches 205; Indels 52; Gaps 14;

QY 34 LLFVIVCLLWLDKSHCAPTWK-----DKTAISENLKSFSEVGEIDAEVKKALTGIQ 87
Db 4 lllfvgl-----twesqvgldgtvsnelqemsnqgskynkeiqnavngvk 54
QY 88 MKIMMERKEHTNLMSTLKKCKREEKQALKLLNEVOHEERLCRESLADSWGECRS 147
Db 55 iktliektneerktllnleaaakkkedalnretsetkkelpgvcnetmmalweackp 114
QY 148 CLENNCRIVT-TCQPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQL 206
Db 115 clkqctcmkfyarvcrgsglvgqlqleeflnqspfyfwmgdr-----idsllendrqq 168
QY 207 TQMEDV----FSQLTVDVNSLNFNRSNVFMQMQEFDQTFQSHFIS-DTDLTEPYFPAP 261
Db 169 thmidvmqdhfrassildefqdrf-----ftrepqdy--hylpfsphrrphff--f 219


```
XX FH Key Location/Qualifiers
FT Peptide 1..22 /label= signal_peptide
FT Protein 23..225 /label= mature_clusterin_alpha_chain
FT Protein 226..446 /label= mature_clusterin_beta_chain
FT Region 74..80 /label= nuclear_localisation_signal
FT Modified-site 86..88 /label= potential_glycosylation_site
FT Modified-site 103..105 /label= potential_glycosylation_site
FT Modified-site 145..147 /label= potential_glycosylation_site
FT Cleavage-site 225..226 /label= potential_glycosylation_site
FT Modified-site /label= interchain_cleavage_site
FT Modified-site 290..292 /note= "leads to the creation of alpha and beta chains"
FT Modified-site 316..318 /label= potential_glycosylation_site
FT Modified-site 353..355 /label= potential_glycosylation_site
FT Modified-site 373..375 /label= potential_glycosylation_site
XX PN WO200034469-A1.
XX PD 15-JUN-2000.
XX PF 10-DEC-1999; 99WO-US29262.
XX PR 11-DEC-1998; 98US-0111856.
XX PA (YUNY ) UNIV NEW YORK STATE RES FOUND.
XX PI Millis AJT;
XX WPI; 2000-431300/37.
XX Clusterin and gp38K-related peptide capable of altering cell migration
XX useful for treating atherosclerosis, cancer and stenosis following
XX vascular trauma or disease
XX Disclosure; Fig 1; 43pp; English.
XX The present sequence is the protein sequence of porcine clusterin.
XX Clusterin (also known as complement cytotoxicity inhibitor, sulfated
XX glycoprotein-2, testosterone repressed prostate message-2, SP-40, 40 and
XX ApoJ) is essential for the migration of vascular smooth muscle cells
XX (VSMC). The gene and protein can, therefore, be used to promote wound
XX healing, angiogenesis and vasculogenesis, in the treatment of stenosis
XX following vascular trauma or disease and to treat atherosclerosis, and
XX antisense sequences can be used to treat cancer, as angiogenesis is
XX vital for tumour survival.
XX Sequence 446 AA;

Query Match 15.2%; Score 400.5; DB 21; Length 446;
Best Local Similarity 24.4%; Pred. No. 6.8e-25;
Matches 118; Conservative 98; Mismatches 196; Indels 71; Gaps 16;

QY 34 LLVFIVCLLWLKDSHCAPTWK-----DKTAISENLKSFVEGEIDAEVKKALTGIK 86
DB 4 lllivglll-----twngpwvlgdkaisdkeigemstegskynkeiknalkev 54
QY 87 QMKTWMERKEHTNLMSTLKKCRKEQALKLLNEVQHELEERLCRESLADSWGECR 146
DB 55 qiktliedsneerkllssleakkkkedaIndrtdetklksggicnetmmalweeck 114
```

```
QY 147 SCLENNOMRIYT-TCQPSWSVSKNKIERFFRKTYIQFLFPFHEDNEKDLPISEKLIKDAQ 205
DB 115 pcikqtcmkfyarvcrrsgslvgqlleeflnqspfyfwingdr-----idslmdndrq 168
QY 206 LTOMEDVFSQLTVDVNSLFRNSNVFRMQQEFDTQFSHFIS-----DTDLTPEY----- 256
DB 169 qshvmdi-----medsfnrnsni-----mdelfqdrfrfpretdtqffspfgsshr 214
QY 257 ---FFPAFSKEPTMKADLEQCWDIPNEFFQLFCNFSVSIYESVSETITKMLKATEDLPKQD 313
DB 215 gslfnpkrsfarnimpflftdl-nyhdmfqpfi-----fdmihgaqqamdahhripyh- 268
QY 314 KAPDHGGLISKMLP--GQDRGLCGELDONLSRCFKFHEKCKQQAHLSEDC-----PDVPA 367
DB 269 -fpeag-----vpensndravckeirhnstgclrmkdqckecreilsvdcsasnsqmq 321
QY 368 LHTELDEAIRLVNVSNOQYQIQLQMTKRKHELTAYLVKMRGQFGWVSELANQAPETETI 427
DB 322 lrqelytslqmaekfskiyqqlqsgyqkmlntssllkqineqfswsqianltqnddry 381
QY 428 FNSIQVVPRIHEGNISKQDETMTMTDLSILPSSNFTLPILESSESSNFYGVVAKALQH 487
DB 382 ylvqttv-nshgsdpsvpsgltkvvklfdsyptiliipqevs--dpkfmnetvaeaalqq 438
QY 488 FKE 490
DB 439 yrq 441

RESULT 15
AAE03764
ID AAE03764 standard; Protein; 363 AA.
AC AAE03764;
XX
DT 07-AUG-2001 (first entry)
XX
DE Human gene 1 encoded secreted protein HOF0C33, SEQ ID NO:34.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiotensin disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW endocrine disorder; infection; wound healing; vulnery;
KW cell culture; chemotaxis; food additive;
KW binding partner identification.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..21 /label= signal_peptide
FT Protein 22..363 /note= "Mature secreted protein"
XX
XX WO200132837-A1.
XX
XX 10-MAY-2001.
XX
XX 17-OCT-2000; 2000WO-US28664.
XX
XX 02-NOV-1999; 99US-0163085.
XX 17-DEC-1999; 99US-0172411.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Fiscella M, Wei P, Lafleur DW, Olsen HS, Baker K, Ebner R;
XX Komatsoulis G, Rosen CA, Ruben SM, Duan RD, Young PE, Florence KA;
XX Moore PA, Birse CE, Ni J, Soppet DR, Shi Y;
```

XX

DR WPI: 2001-328782/34.
DR N-PSDB; AAD08191.

XX

PT Novel human secreted proteins and nucleic acids for diagnosing,
PT preventing and treating neurological, cardiovascular, infectious,
PT autoimmune, gastrointestinal, bone disorders, cancer, particularly
PT ovarian cancer

XX

PS Claim 11; Page 386-387; 421pp; English.

XX

CC AAD08191-AAD08213 represent cDNAs corresponding to 19 human secreted
CC protein genes and AAE03764-AAE03786 represent the proteins they encode.
CC AAE03787-AAE03800 represent human secreted protein fragments or variants.
CC The genes and their secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 19 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention.

XX

SQ Sequence 363 AA;

Query Match 12.2%; Score 320; DB 22; Length 363;

Best Local Similarity 22.6%; Pred. No. 2.6e-18;

Matches 86; Conservative 80; Mismatches 126; Indels 88; Gaps 12;

.QY 34 LLVFI VCLLWLKDSHCAPTWKDKTATSENLSKSFSEVGEIDAD-----EEVKALGTGIKQ 87
Db 4 llllvglll-----twengrvgldqnvdsdtelqemstegskynrelnknlxgvkq 54
QY 88 MKIMMERKEHTNLMSTLKCKREEKQEA KLLNEVQEHLEEEERLCRESLADSWGECRS 147
Db 55 llllqteqneerkslltnleakkkkedalndtkdsemkllkasggvcndtmmlweeckp 114
QY 148 CLENENCMRIYT-TCQPSWSVKNKIERFRKTYQFLFPFHEDNEKDLPISEKLIKDAQL 206
Db 115 ctkqcmkfyarvcrsstglvgqvveeflnqspfyfwngdr-----idlallendrqg 168
QY 207 TQMEDVFSQLTVDVNSLFRNFRNQMQEFDQTFQSHFISDITDLETPYFFPAFSK--- 263
Db 169 thaldv-----mqdsforassi-----mdelfqdrfft-raqqpfhfspsfqr 213
QY 264 -----EPMTKADLEQCWDIPNFFQFCNFSVSIYSVSETITKM 302
Db 214 rpfffnikhrfarnimpfgyqpl-----nfhdmfqpfddmlhga-qqamdv 260
QY 303 LKAIEDLPKQDKAPDHGGLISKNLPGQDRGLCGELDONLSRCFKEKQCKQAHLSDEC 362
Db 261 lhrphfpmfesteedn-----qdgavckehrhnstgclknkdqckecreilsvdc 310
QY 363 ----PDVPALHTELDEAIRL 378

Db 311 ssnpqavqlrgelnslql 330

Search completed: July 2, 2002, 11:51:19
Job time: 397 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:48:57 ; Search time 23.96 Seconds
(without alignments)
504.619 Million cell updates/sec

Title: US-09-722-544A-2MOD
Perfect score: 2632
Sequence: 1 MKIAEKNEGSRWQLHW.....FTGVYVAKALQHFKEHFKTW 495

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2628	99.8	495	4	US-09-268-992-2
2	2460	93.5	466	4	US-09-268-992-64
3	2346	89.1	446	4	US-09-268-992-51
4	2331	88.6	477	4	US-09-268-992-4
5	1751.5	66.5	465	4	US-09-268-992-49
6	1635	62.1	466	4	US-09-268-992-39
7	1554	59.0	450	4	US-09-268-992-41
8	1277	48.5	374	4	US-09-268-992-43
9	1272.5	48.3	373	4	US-09-268-992-45
10	921	35.0	208	4	US-09-268-992-67
11	149	5.7	1388	2	US-08-685-576-1
12	145	5.5	1388	2	US-08-685-576-4
13	131	5.0	816	2	US-08-533-306A-6
14	131	5.0	816	2	US-08-742-923A-6
15	130.5	5.0	2462	1	US-08-328-254-6
16	130	4.9	3248	5	US-08-353-700-1
17	130	4.9	3248	5	PCT-US95-16216-1
18	129	4.9	885	2	US-08-533-306A-4
19	129	4.9	885	2	US-08-742-923A-4
20	126.5	4.8	1354	3	US-08-685-871-2
21	123	4.7	1618	1	US-07-853-913-4
22	122	4.6	1886	4	US-08-938-105-3
23	121	4.6	828	2	US-08-993-228-21
24	119.5	4.5	435	2	US-08-531-439B-4
25	119.5	4.5	1786	4	US-08-973-462-8
26	119.5	4.5	3111	2	US-08-460-309-4
27	119.5	4.5	3111	2	US-08-125-077-4

28 117 4.4 1098 4 US-08-923-992A-8 Sequence 8, Appli
29 117 4.4 1104 4 US-08-923-992A-4 Sequence 4, Appli
30 117 4.4 1939 4 US-09-310-187A-1 Sequence 1, Appli
31 117 4.4 2052 3 US-09-045-201A-2 Sequence 2, Appli
32 116.5 4.4 1057 4 US-09-541-782-10 Sequence 10, Appli
33 116 4.4 1164 4 US-08-923-992A-2 Sequence 2, Appli
34 116 4.4 2154 2 US-08-841-349-4 Sequence 4, Appli
35 115.5 4.4 452 2 US-08-686-599A-18 Sequence 18, Appli
36 115.5 4.4 493 2 US-08-686-599A-5 Sequence 5, Appli
37 115.5 4.4 493 2 US-08-686-599A-16 Sequence 16, Appli
38 114 4.3 1128 4 US-08-923-992A-6 Sequence 6, Appli
39 113.5 4.3 467 2 US-08-686-599A-17 Sequence 17, Appli
40 111.5 4.2 1066 4 US-09-541-782-8 Sequence 8, Appli
41 109 4.1 976 4 US-09-104-324B-4 Sequence 4, Appli
42 109 4.1 1093 5 PCT-US93-03077-1 Sequence 1, Appli
43 108 4.1 2285 4 US-09-308-375-2 Sequence 2, Appli
44 107 4.1 1164 4 US-08-923-992A-10 Sequence 10, Appli
45 107 4.1 1497 1 US-08-623-679-7 Sequence 7, Appli

ALIGNMENTS

RESULT 1

US-09-268-992-2
; Sequence 2, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Frelmer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-2

Query Match 99.8%; Score 2628; DB 4; Length 495;

Best Local Similarity 99.8%; Pred. No. 7.1e-245;

Matches 494; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKIAEKNEGSRWQLHWGDIANNNGMKPPLLVFTVCLLWLDKSHCAPTWKDKTAIS 60
DB 1 MKIAEKNEGSRWQLHWGDIANNNGMKPPLLVFTVCLLWLDKSHCAPTWKDKTAIS 60
QY 61 ENLKSFSVEGIDAEVVKKALTGTIKQKIMMERKEKHTNLMSTLKKCKREEQKALKL 120
DB 61 ENLKSFSVEGIDAEVVKKALTGTIKQKIMMERKEKHTNLMSTLKKCKREEQKALKL 120
QY 121 NEVOEHLEEEERLCRESLADSWGECRSCLENNCMRIYTCOPSSSVKNKIERFRKIYQ 180
DB 121 NEVOEHLEEEERLCRESLADSWGECRSCLENNCMRIYTCOPSSSVKNKIERFRKIYQ 180
QY 181 FLFFPHEDNEKDLPTSEKLEIEKDAQLTQMEDVFSQLTVDVNSLNRSNFVRQMOQEFQ 240
DB 181 FLFFPHEDNEKDLPTSEKLEIEKDAQLTQMEDVFSQLTVDVNSLNRSNFVRQMOQEFQ 240
QY 241 TFQSHFISDTDLTEPFYFFPAFSKEPMTKADLEQCQWDIPNFFQFCNFVSIVSVSETIT 300

Db 241 TFSHLEISDLDTEPFFAFSKEPMTKADLEQCDWLPNFFQLFCNFSVIYESVSETIT 300
QY 301 KMLKAIEDLPKQDAPDHGGLISKMLPGQDRGLCGELDNLSRCFKFHEKQCKQCAHLSE 360
Db 301 KMLKAIEDLPKQDAPDHGGLISKMLPGQDRGLCGELDNLSRCFKFHEKQCKQCAHLSE 360
QY 361 DCDPVPALHTELDEAIRLVNVSNOQYQGIQLQMTKRKHELTAYLVEKMRGQFGWVSELANO 420
Db 361 DCDPVPALHTELDEAIRLVNVSNOQYQGIQLQMTKRKHELTAYLVEKMRGQFGWVSELANO 420
QY 421 APTEIIFNSIQVVPRIHEGNISKQDETMTDLSILPSSNFTLKIPLESASSNFIGYV 480
Db 421 APTEIIFNSIQVVPRIHEGNISKQDETMTDLSILPSSNFTLKIPLESASSNFIGYV 480
QY 481 VAKALQHFEHFKTW 495
Db 481 VAKALQHFEHFKTW 495

RESULT 2
US-09-268-992-64
; Sequence 64, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-64

Query Match 93.5%; Score 2460; DB 4; Length 466;
Best Local Similarity 99.8%; Pred. No. 1e-228;
Matches 465; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 30 MKPPLLVFVCLLWLKXDSHCAPTWKDKTAISENLKSFSEVGEIDADEEVKKALTGIKQMK 89
Db 1 MKPPLLVFVCLLWLKXDSHCAPTWKDKTAISENLKSFSEVGEIDADEEVKKALTGIKQMK 60
QY 90 IMMERKEHTNLMSTLKKREKQBALKLLNEVQEHLEERLCRESLADSWGECRSL 149
Db 61 IMMERKEHTNLMSTLKKREKQBALKLLNEVQEHLEERLCRESLADSWGECRSL 120
QY 150 ENNCMIYITTCQPSWSVKNKTERFFRKIYQFLFPFHEDNEKDLPISEKLIKEDAQLTOM 209
Db 121 ENNCMIYITTCQPSWSVKNKTERFFRKIYQFLFPFHEDNEKDLPISEKLIKEDAQLTOM 180
QY 210 EDVFSOLTVDVNSLNRNFVFMQOEEDQTFQSHFISDITDTEPYFFPAFSKEPMTKA 269
Db 181 EDVFSOLTVDVNSLNRNFVFMQOEEDQTFQSHFISDITDTEPYFFPAFSKEPMTKA 240
QY 270 DLEQCWDIPNFQFCNFSVYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 329
Db 241 DLEQCWDIPNFQFCNFSVYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 300

QY 330 DRGLCGELDONLSRSCFKFHEKQCKQCAHLSEDCPDVPALHTELDEAIRLVNVSNOQYQGI 389
Db 301 DRGLCGELDONLSRSCFKFHEKQCKQCAHLSEDCPDVPALHTELDEAIRLVNVSNOQYQGI 360
QY 390 LQWTRKHLEDYAYLVEKMRGQFGWSELANOQAPETELIENSIOVVPRIHEGNISKQDETMT 449
Db 361 LQWTRKHLEDYAYLVEKMRGQFGWSELANOQAPETELIENSIOVVPRIHEGNISKQDETMT 420
QY 450 MTDLSILPSSNFTLKIPLESASSNFIGYVYVAKALQHFEHFKTW 495
Db 421 MTDLSILPSSNFTLKIPLESASSNFIGYVYVAKALQHFEHFKTW 466

RESULT 3
US-09-268-992-51
; Sequence 51, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 51
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-51

Query Match 89.1%; Score 2346; DB 4; Length 446;
Best Local Similarity 99.8%; Pred. No. 9.2e-218;
Matches 445; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 50 APTWKDKTAISENLKSFSEVGEIDADEEVKKALTGIKQMKIMMERKEHTNLMSTLKKC 109
Db 1 APTWKDKTAISENLKSFSEVGEIDADEEVKKALTGIKQMKIMMERKEHTNLMSTLKKC 60
QY 110 REEQBALKLLNEVQEHLEERLCRESLADSWGECRSCLENNCMRIYITTCQPSWSVKN 169
Db 61 REEQBALKLLNEVQEHLEERLCRESLADSWGECRSCLENNCMRIYITTCQPSWSVKN 120
QY 170 KIERFRKIYQFLFPFHEDNEKDLPISEKLIKEDAQLTOMEDVFSOLTVDVNSLNRNSEN 229
Db 121 KIERFRKIYQFLFPFHEDNEKDLPISEKLIKEDAQLTOMEDVFSOLTVDVNSLNRNSEN 180
QY 230 VFQMOQOEEDQTFQSHFISDITDTEPYFFPAFSKEPMTKADLEQCWDIPNFQFCNFSV 289
Db 181 VFQMOQOEEDQTFQSHFISDITDTEPYFFPAFSKEPMTKADLEQCWDIPNFQFCNFSV 240
QY 290 SIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQDRGLCGELDONLSRSCFKPHE 349
Db 241 SIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQDRGLCGELDONLSRSCFKPHE 300
QY 350 KQCKQCAHLSEDCPDVPALHTELDEAIRLVNVSNOQYQGILOWTRKHLEDYAYLVEKMRG 409
Db 301 KQCKQCAHLSEDCPDVPALHTELDEAIRLVNVSNOQYQGILOWTRKHLEDYAYLVEKMRG 360
QY 410 QFGWVSELANOQAPETELIENSIOVVPRIHEGNISKQDETMTDLSILPSSNFTLKIPLEE 469
Db 361 QFGWVSELANOQAPETELIENSIOVVPRIHEGNISKQDETMTDLSILPSSNFTLKIPLEE 420

QY 470 SAESSNFIGYVYVAKALQHFKEHFKTW 495
|||||
Db 421 SAESSNFIGYVYVAKALQHFKEHFKTW 446

RESULT 4
US-09-268-992-4
; Sequence 4, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-4

Query Match 88.6%; Score 2331; DB 4; Length 477;
Best Local Similarity 91.0%; Pred. No. 2.9e-216;
Matches 44; Conservative 3; Mismatches 1; Indels 40; Gaps 2;
QY 25 NSNGMKPPLLVFVCLLWKDSCAPTWKDTAISNLKSFSEVGEIDAEVKKALTG 84
Db 7 SNSGMMKPLLVFVCLLWKDSCAPTWKDTAISNLKSFSEVGEIDAEVKKALTG 66
QY 85 IKQKIMMERKEK-----EHTNLMSTLKKCKREKQKALKLLNEVQ 124
Db 67 IKQKIMMERKEKANQAPETEIIFNSIQVVPRIEHTNLMSTLKKCKREKQKALKLLNEVQ 126
QY 125 EHLKEERLCRESLADSWGECRSCLENNCMRIYTTQCPSSVSKNKIERFRKIYQFLFP 184
Db 127 EHLKEERLCRESLADSWGECRSCLENNCMRIYTTQCPSSVSKNKIERFRKIYQFLFP 186
QY 185 PHEDNEKDLPISEKLEBKDAQLTQMEDVFSQLTVDVNSLFRNSFNFRMQQEDDTFQS 244
Db 187 PHEDNEKDLPISEKLEBKDAQLTQMEDVFSQLTVDVNSLFRNSFNFRMQQEDDTFQS 246
QY 245 HFISDLDLTPYFPFAPSKPEMTKADLEQCDWIPNFFQFCNFSVSVSETITKMLK 304
Db 247 HFISDLDLTPYFPFAPSKPEMTKADLEQCDWIPNFFQFCNFSVSVSETITKMLK 306
QY 305 AIEDLPQDKAPDHGGLISKMLPGDRLGELDONLSRCFKHEKCKQKQAHLSDCPD 364
Db 307 AIEDLPQDKAPDHGGLISKMLPGDRLGELDONLSRCFKHEKCKQKQAHLSDCPD 366
QY 365 VPALHTELDEAIRLVNSNOQYQILQWTRKHLDTAYLVKMRGQFGWVSELANQAPET 424
Db 367 VPALHTELDEAIRLVNSNOQYQILQWTRKHLDTAYLVKMRGQFGWVSELANQAPET 424
QY 425 EIIFNSIQVVPRIEHNISQKDETMVTDLSILPSSNFTLKIPIEESAESSNFIGYVYVAKA 484
Db 420 -----HEGNISQKDETMVTDLSILPSSNFTLKIPIEESAESSNFIGYVYVAKA 466
QY 485 LQHFKEHFKTW 495
|||||

Db 467 LQHFKEHFKTW 477

RESULT 5
US-09-268-992-49
; Sequence 49, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 49
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Bos sp.
US-09-268-992-49

Query Match 66.5%; Score 1751.5; DB 4; Length 465;
Best Local Similarity 70.8%; Pred. No. 2e-160;
Matches 329; Conservative 59; Mismatches 77; Indels 1; Gaps 1;
QY 30 MKPPLLVFVCLLWKDSCAPTWKDTAISNLKSFSEVGEIDAEVKKALTGIKOMK 89
Db 1 MKPPLLVFVCLLWKDSCAPTWKDTAISNLKSFSEVGEIDAEVKKALTGIKOMK 60
QY 90 IMMERKEKHTNLMSTLKKCKREKQKALKLLNEVQEHLEERLCRESLADSWGECRSC 149
Db 61 ILMEREEHSHKMLRTLKKCKREKQKALKLLNEVQEHLEERLCRESLADSWGECRSC 120
QY 150 ENNCMRIYTTQCPSSVSKNKIERFRKIYQFLPPPHEDNEKDLPISEKLEBKDAQLTOM 209
Db 121 ESDCMRYTTQCPSSVSKNKIERFRKIYQFLPPPHEDNEKDLPISEKLEBKDAQLTOM 180
QY 210 EDVFSQLTVDVNSLFRNSFNFRMQQEDDTFQSHFISDLDLTPYFPFAPSKPEMTKA 269
Db 181 ENVFSQLTVDVNSLFRNSFNFRMQQEDDTFQSHFISDLDLTPYFPFAPSKPEMTKA 240
QY 270 DLQECWDIPNFFQFCNFSVSVSETITKMLKAIEDLPKODKAPDHGGLISKMLPGQ 329
Db 241 HPMQSDWIPNFFQFCNFSVSVSETITKMLKAIEDLPKODKAPDHGGLISKMLPGQ 300
QY 330 DRGLCGELDONLSRCFKHEKCKQKQAHLSDCPDVPAHTEDEAIRLVNSNOQYQGI 389
Db 301 GRGLCGEPGNSSECLQFHARCKQKQDYLWADCPANPELYTKADEALELVNISNQTAQV 360
QY 390 LQWTRKHLDTAYLVKMRGQFGWVSELANQAPETEIIFNSIQVVPRIEHNISQKDETM 449
Db 361 LQWTRKHLDTAYLVKMRGQFGWVSELANQAPETEIIFNSIQVVPRIEHNISQKDETM 420
QY 450 MTDLSILPSSNFTLKIPIEESAESSNFIGYVYVAKALQHFKEHFKTW 495
Db 421 I-DISILPSSNFTLKIPIEESAESSNFIGYVYVAKALQHFKEHFKTW 465

RESULT 6
US-09-268-992-39
; Sequence 39, Application US/09268992
; Patent No. 6342351

```
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER FILING DATE: 1998-10-28
; EARLIER FILING DATE: 1998-10-28
; EARLIER FILING DATE: 1998-06-05
; EARLIER FILING DATE: 1998-06-05
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Cavia sp.
US-09-268-992-39

Query Match      62.1%; Score 1635; DB 4; Length 466;
Best Local Similarity 66.5%; Pred. No. 3.4e-149;
Matches 311; Conservative 64; Mismatches 89; Indels 4; Gaps 3;

QY 30 MKPPLLVTIVCLLWLDKSHCAPTKDKTAISENLKSFSEVGEIDADVEVKKALTGKQMK 89
DB 1 MKPLPMLFPVCLLWLDKCHCAPTKDKTAISENANFSSEAGEIDVDGEVKIALIGIKQMK 60
QY 90 IMMERKEHTNLMSTLKKCKREEQBALKLNEVQHLNEERLCSLADSGEGRSCL 149
DB 61 IMMERREEHSLMKTLKKCKEKEQBALKLNEVHEHLEESLCOVSLADSWDECRACL 120
QY 150 ENNCRIYTCOPSSWSVKNKIERFERKIQFLFPFHEDNEKDLPISEKLIKEDAOQTOM 209
DB 121 ESNCRMFDITCOPAWSSVKNM-----ENDRSGPVSKGVTEEDAQVSHI 179
QY 210 EDVFSQLTVDVNSLNFNRSENFVQMOQEFDDQTFQSHFISDTDLTEPYFFPAFSPKPTKA 269
DB 180 EHVSQLSADVTSLFNRSLYVFKLRREFDQAFQSVFTSGTDVTEPFPPSLSKEPAYRA 239
QY 270 DLQCDWIPNFFQCNFSVSIYESVETITKMLKAIEDLPKODKAPDHGGLISKMLPGQ 329
DB 240 DAEPSWAIPNVFQLLCNLSFSVYQSVSEKLTITTLRATEDPPKQKDSNOGGPISKILPEQ 299
QY 330 DRGLCGELDONLSRCFKFHEKQKCOAHLSDEDCPDVPALHTEDEAIRLVNYSNOQYGOI 389
DB 300 DRGSDGKLGONLSDCNVFRKRCQKQDYLSDDCPNVPELYRELNEALRLVSRNQYDQV 359
QY 390 LQMTKRHLEDYALVEKMRQFGWVSELANQAPETEIFNSIQVVPRI--HEGNISKQDE 447
DB 360 VQMTQYHLEDITLLMEKMRQFGWVSELAYQSPGAEDIFNPVKVMVALSAHEGNSDDQD 419
QY 448 TMMTDLSILPSSNFTLKIPLSEASNSNFIQYVYVAKALQHFKEHFKTW 495
DB 420 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVQLQHFKEHFKTW 466

RESULT 7
US-09-268-992-41
; Sequence 41, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER FILING DATE: 1998-06-05
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Cavia sp.
US-09-268-992-39
```

```
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Cavia sp.
US-09-268-992-41

Query Match      59.0%; Score 1554; DB 4; Length 450;
Best Local Similarity 63.7%; Pred. No. 2e-141;
Matches 298; Conservative 63; Mismatches 87; Indels 20; Gaps 3;

QY 30 MKPPLLVTIVCLLWLDKSHCAPTKDKTAISENLKSFSEVGEIDADVEVKKALTGKQMK 89
DB 1 MKPLPMLFPVCLLWLDKCHCAPTKDKTAISENANFSSEAGEIDVDGEVKIALIGIKQMK 60
QY 90 IMMERKEHTNLMSTLKKCKREEQBALKLNEVQHLNEERLCSLADSGEGRSCL 149
DB 61 IMMERREEHSLMKTLKKCKEKEQBALKLNEVHEHLEESLCOVSLADSWDECRACL 120
QY 150 ENNCRIYTCOPSSWSVKNKIERFERKIQFLFPFHEDNEKDLPISEKLIKEDAOQTOM 209
DB 121 ESNCRMFDITCOPAWSSVKNM-----ENDRSGPVSKGVTEEDAQVSHI 163
QY 210 EDVFSQLTVDVNSLNFNRSENFVQMOQEFDDQTFQSHFISDTDLTEPYFFPAFSPKPTKA 269
DB 164 EHVSQLSADVTSLFNRSLYVFKLRREFDQAFQSVFTSGTDVTEPFPPSLSKEPAYRA 223
QY 270 DLQCDWIPNFFQCNFSVSIYESVETITKMLKAIEDLPKODKAPDHGGLISKMLPGQ 329
DB 224 DAEPSWAIPNVFQLLCNLSFSVYQSVSEKLTITTLRATEDPPKQKDSNOGGPISKILPEQ 283
QY 330 DRGLCGELDONLSRCFKFHEKQKCOAHLSDEDCPDVPALHTEDEAIRLVNYSNOQYGOI 389
DB 284 DRGSDGKLGONLSDCNVFRKRCQKQDYLSDDCPNVPELYRELNEALRLVSRNQYDQV 343
QY 390 LQMTKRHLEDYALVEKMRQFGWVSELANQAPETEIFNSIQVVPRI--HEGNISKQDE 447
DB 344 VQMTQYHLEDITLLMEKMRQFGWVSELAYQSPGAEDIFNPVKVMVALSAHEGNSDDQD 403
QY 448 TMMTDLSILPSSNFTLKIPLSEASNSNFIQYVYVAKALQHFKEHFKTW 495
DB 404 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVQLQHFKEHFKTW 450

RESULT 8
US-09-268-992-43
; Sequence 43, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER FILING DATE: 1998-06-05
```

; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 43
; LENGTH: 374
; TYPE: PR1
; ORGANISM: Cavia sp.
US-09-268-992-43

Query Match 48.5%; Score 1277; DB 4; Length 374;
Best Local Similarity 54.5%; Pred. No. 7.7e-115;
Matches 255; Conservative 44; Mismatches 73; Indels 96; Gaps 3;
QY 30 MKPPLLFIYVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDADEVYKALTGIKQMK 89
DB 1 MKPLLPFPVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDADEVYKALTGIKQMK 60
QY 90 IMMERKEHTNLMSTLKKCKREEKQKALKLLNEVQHELEERLCRESLADSWGECRSL 149
DB 61 IMMERREEHSLKMTLKKCKEKEQKALKLLNEVHEHELEESLCQVSLADSWDECRACL 120
QY 150 ENCMRIYTCQPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLEKDAQLTQM 209
DB 121 ESNCRFDTTCPAWSVKN----- 140
QY 210 EDVFSQTLVDVNSLNFNSFVFRMQOEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 269
DB 141 -----NEPAYRA 147
QY 270 DLEQCWDIPNFQFCNFSYISVSSETITKMLKAIEDLPKQKADPHGGLISKMLPGQ 329
DB 148 DAEPWAIPNVFOLLNLSFVSQVSEKLTTLRATEDPPKQKDSNOGPGISKILPEQ 207
QY 330 DRGLCGELDONLSRCFKFHEKQKQAHLSDECDPVPALHTDELDEAIRLVNVSNOQYQGI 389
DB 208 DRGSDGKLGONLSDCVNFRKRCQKQDYLSDDCPNVPPELYRELNEALRLVSRNSNOQYDQV 267
QY 390 LQMTKHELTAYLVEKMRGQFGWVSELANOAPETIIFNSIOVVPRI--HEGNISKODE 447
DB 268 VQMTQYHLEDTLLMEKREQFGWVSELAYQSPGAEDIFNPVKVVALSHEGNSDQDD 327
QY 448 TMMDLSILPSSNFTLKIPLESSESSNFIYGVYAKALQHFKEHFTW 495
DB 328 TVWPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVLQHFKEHFTW 374

RESULT 9
US-09-268-992-45
; Sequence 45, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45
; LENGTH: 373
; TYPE: PR1

; ORGANISM: Cavia sp.
US-09-268-992-45

Query Match 48.3%; Score 1272.5; DB 4; Length 373;
Best Local Similarity 54.5%; Pred. No. 2.1e-114;
Matches 255; Conservative 45; Mismatches 71; Indels 97; Gaps 4;
QY 30 MKPPLLFIYVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDADEVYKALTGIKQMK 89
DB 1 MKPLLPFPVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDADEVYKALTGIKQMK 60
QY 90 IMMERKEHTNLMSTLKKCKREEKQKALKLLNEVQHELEERLCRESLADSWGECRSL 149
DB 61 IMMERREEHSLKMTLKKCKEKEQKALKLLNEVHEHELEESLCQVSLADSWDECRACL 120
QY 150 ENCMRIYTCQPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLEKDAQLTQM 209
DB 121 ESNCRFDTTCPAWSVKN----- 140
QY 210 EDVFSQTLVDVNSLNFNSFVFRMQOEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 269
DB 141 -----MPAYRA 146
QY 270 DLEQCWDIPNFQFCNFSYISVSSETITKMLKAIEDLPKQKADPHGGLISKMLPGQ 329
DB 147 DAEPWAIPNVFOLLNLSFVSQVSEKLTTLRATEDPPKQKDSNOGPGISKILPEQ 206
QY 330 DRGLCGELDONLSRCFKFHEKQKQAHLSDECDPVPALHTDELDEAIRLVNVSNOQYQGI 389
DB 207 DRGSDGKLGONLSDCVNFRKRCQKQDYLSDDCPNVPPELYRELNEALRLVSRNSNOQYDQV 266
QY 390 LQMTKHELTAYLVEKMRGQFGWVSELANOAPETIIFNSIOVVPRI--HEGNISKODE 447
DB 267 VQMTQYHLEDTLLMEKREQFGWVSELAYQSPGAEDIFNPVKVVALSHEGNSDQDD 326
QY 448 TMMDLSILPSSNFTLKIPLESSESSNFIYGVYAKALQHFKEHFTW 495
DB 327 TVWPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVLQHFKEHFTW 373

RESULT 10
US-09-268-992-67
; Sequence 67, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 208
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-268-992-67

Query Match 35.0%; Score 921; DB 4; Length 208;
Best Local Similarity 99.4%; Pred. No. 6.5e-81;
Matches 170; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MK1AEKNEGPRSWQLHWGDIANNNGNNKPPLLVFIVCLLWLKDSHCAPTWKOKTAS	60
Db	33	MK1AEKNEGPRSWQLHWGDIANNNGNNKPPLLVFIVCLLWLKDSHCAPTWKOKTAS	92
Qy	61	ENLKSSEVGEIDADAEVVKALGTGIOMKIMMERKEKHTNLMSTLUKKRESEKQALKLL	120
Db	93	ENLKSSEVGEIDADAEVVKALGTGIOMKIMMERKEKHTNLMSTLUKKRESEKQALKLL	152
Qy	121	NEVQEHLEERLCRESLADSGECRSCLENNCRMRYTTTCQPSWSVYKNI	171
Db	153	NEVQEHLEERLCRESLADSGECRSCLENNCRMRYTTTCQPSWSVYKNI	203

```

RESULT 11
US-08-685-576-1
; Sequence 1, Application US/0868576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Kozo
; APPLICANT: Iwamatsu, Akihito
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uak1
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996

```

Query Match 5.7%; Score 149; DB 2; Length 1388;
Best Local Similarity 16.7%; Pred. No. 3e-05;
Matches 95; Conservative 96; Mismatches 166; Indels 212; Gaps 20;
QY 54 KDKTAISENLKSFSEVEIGDADEVKALGVIQMKI-----MMERKKEKHTNLM 104

Db	780	QKQVLDNEDVRNLT	---	LKIEQETQKRCLTQNDLKMQTQOVNTLKMKSEQLKQKNNHLL	835		
Qy	105	TLK	-----	KCREBKQALKLLNNEVQHELEEE	---	RLCR	135
Db	836	EMKMLEKQNAELRKERQDAGQMKELQDLEAEQYFSTLYKTVQRELKECEKTKLCK	895				
Qy	136	E	-----	SLADSWGCRSCLENNCMRIYTTQPSWSSVKNKIERFRFKIYQFLFPFHEDN	189		
Db	896	ELQOKQELQDE	-----	-----	RDSLAAQLEITLTKADSEQLARSIAEQYSDLEKE	942	
Qy	190	E	-KDLPISE-	-----	KLIEKDAOLTOMEDVFESOLTVDVNSLNRSPNFRMQQEFQD	240	
Db	943	KIMKELEIKEMARHKQELTEKDATIASLEETNRTLTSVANLANEKEELNNKLEAQEQ	1002				
Qy	241	TFQSHFISDTDLTEPYFFPAFSKEPMKADLEQCQWDIPNFFQLFCNFSVSIYESVETIT	300				
Db	1003	LSR	-----	LKDEISAAAIAKAFQKLLTERTLK	-----	TQAVN	1036
Qy	301	KMLKATE	-----	DLPQDKAPDHGGLLSKMLPGQDRGLCGEL	-----	DONLSRCF	345
Db	1037	KLAEIWNKPEVKGNDTVRRKE	-----	-----	ENRKLHMLKSEKRLTQOMI	1082	
Qy	346	KPHEKQKQCAHLSEDCDPVPALHTDELDEAIRLVNVSQQYQGLIQMTRKHLE	---	---	---	---	398
Db	1083	KYQKELNEMQAQIAEB	-----	SQIRIELQMTLDSKSDIEQLRSQLQALHGLDSSSTGSG	1138		
Qy	399	-----	DTAYLVKEKMRG	-----	QFGWSEL	-----	417
Db	1139	PGTEADDDGFPESRUEGLWSLVPVRNNTKFGWKYKIVTSSKKILFYDSEODKEQSNPYM	1198				
Qy	418	-----	ANOAPETEIFNSIQVVPRI	-----	HEGNISKODETMTWTLDSILPSSNFT	462	
Db	1199	VLDIDKLEHVRPVTQDVRADAKELPRIFQILLYANEGESKKEQE	-----	-----	-----	-----	1243
Qy	463	LK1PLEESAESSNFI	---	GYVVAKALQHF	488		
Db	1244	--FPVSPVGEKSNYICHKHGHEEPIPLYHF	1270				

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```



```

; FILING DATE: 05-JAN-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-685-576-4

```

```

Query Match      5.5%; Score 145; DB 2; Length 1388;
Best Local Similarity 17.0%; Pred. No. 7.4e-05;
Matches 95; Conservative 96; Mismatches 176; Indels 192; Gaps 20;

QY 54 KDKTAISENLKSFSEVGIDEADEVKKALTGKQMKI-----MMERKEKEHTNLMS 104
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 780 KQKDVNLNEDVRNLT-----LKIEQETQKRCLTQNDLKMOTQOVNTLKMSEKQLKQENHLM 835
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 105 TLK-----KCREQKQKALKNEVOEHLEEE-----RLCRESLADSWGEC----- 145
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 836 EMKMNLEKQNELKRDADQKQKQLEAQYFTLYKTQVRELKECECKTLGK 895
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 146 ---RSCLENNCMRIYTTQCPSSVSKKIERFFRKIYFLFPFHEDNE--KDLPISE--- 197
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 896 ELQKKQKQELQDERSLAQLEITLKADSEQLARSIAEQYSDLEKEKIMKEIKEMMA 955
QY 198 ---KLIKDAQLTQMEDVFSQLVDVNSLFRS---FNVFPMQOQEFQFQSHFISDT 250
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 956 RHQBELTEKDATIASLEETNRTLSADVANLANEKEELNKLKDVQQLSR-----LKDE 1009
QY 251 DLTEPYFFPAPFSKPEMTKADLEQCDWIPNFFQFCNFVSIVSEVETITKMLKAIE--- 307
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1010 EISAAAIKAQFEKQLLTERTLK-----TQVANKLAEMNRKE 1046
QY 308 -----DLPKQKADPHGGLISKMLPGDQRLGCGEL-----DQNLSCRCFKFHEKQKQ 355
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1047 PVKRGNDTVRRKEK-----ENRKLHMLKSERKLTQOMIKYQKELNEMQ 1092
QY 356 AHLSEDCPDVPAHTELDIAIRLVNSNQYQGQILQWTRKHE-----DTAY 402
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1093 AQIAEE---SQIRIELQMTLDSKDSIEQLRSQALHGLDSSIGSGPGDAEADGGF 1148
QY 403 LVKMRG-----QFGWSEL-----A 418
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1149 PESRLEGLWLSLPVRNNTKFGWVKYIVSSKKILFYDSQDKQSNPYVAVDIDKLFV 1208
QY 419 NQAPETELIFNSIQVPR-----HEGNTSKQDETMTDLISLPSSNFTIKIPLERSAE 472
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1209 RPTQTDTYRADAKEIPRIFOILYANEGESKKEQ-----FPVEPVGE 1251
QY 473 SSNFI---GYVVAKALQHF 488
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1252 KSNYCHKGHEFIPTLYHF 1270

```

```

RESULT 13
US-08-533-306A-6
; Sequence 6, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.

```

```

; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,306A
; FILING DATE: September 25, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, DeAnn F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869COB
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-533-306A-6

```

```

Query Match      5.0%; Score 131; DB 2; Length 816;
Best Local Similarity 19.0%; Pred. No. 0.00073;
Matches 96; Conservative 81; Mismatches 177; Indels 150; Gaps 20;

QY 4 KAEKNEGSPSRWQLHWGDIANNNGN--MKPPLLVFIVCLLWLDKSHCAPTKWDKTAISE 61
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 NLKSFSEVGIDEADEE-----VVKALTGKIMKIMMERKEKEHTNLM-----STLKK 108
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 -LQRLDGMGCLFDEERAQOQEDALAQQAFEARRTREFEDRDRSHREEMENEVEVTGM 173
QY 109 CREKQKQKALKL-----LNEVOEHLEERELC-----RESLADSWGEC 145
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 146 RSCLENNCMRIYTTQCPSSVSKKIERFFRKIYFLFPFHEDNEKDL-----ISEKLI 200
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 MEA-KONLERHISTNLQSLSDSKKQLQDFASTVEAL-----EEGKRFRQKEIENLQYQE 287
QY 201 EKDAQLTQMEDVFSQLTVDVNSLF-----NRSFNVPROMQOQFDO-----TFQSHFIS 248
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 288 EKAATAVDKLEKTKNRLQQLQELDLVLDLNQRLVSNLEKKQKQKFDQLLAEKRNISKYAD 347
QY 249 DTDLTPYFFPAPFSKPEMTKADLEQCDWIPNFFQFCNFVSIVSEVETITKMLKAIED 308
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 348 ERDRAE-----AEAREKETKA-----LSLARAEEALEAKEE 379
QY 309 LPKQKADPHGGLISKMLPGDQRLGCGELDQNLSCRCFKFHEKQKQAHLSDCPDVPA 368
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 380 LER-----TNKMLKAEMEDLVSSKDDVGNVHELEKSKRALETQMEEMKTQLEEL 429
QY 369 HTL-----DEAIRL-VNVSNQYQGQILQWTRKHELTAYLVKMRGQFGWSELANQAPE 423
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 430 EDELQASEDAKURLEVN-----MOALKGQF-----ERDLQARD 462

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:52:07 ; Search time 17.57 seconds
(without alignments)
1090.847 Million cell updates/sec

Title: US-09-722-544A-2MOD

Perfect score: 2632

Sequence: 1 MKIKAEKNEGFSRWQLHW.....FIGYVVAQALQHFKEHFKTW 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	434	16.5	449	1	CLUS_HUMAN
2	418	15.9	451	1	CLUS_COTJA
3	413.5	15.7	447	1	CLUS_RABIT
4	410.5	15.6	445	1	CLUS_CANFA
5	403	15.3	448	1	CLUS_MOUSE
6	400.5	15.2	439	1	CLUS_BOVIN
7	400.5	15.2	446	1	CLUS_PIG
8	380	14.4	447	1	CLUS_RAT
9	168	6.4	191	1	CLUS_MESAU
10	151	5.7	1526	1	MYS2-SCHPO
11	145.5	5.5	594	1	CIK1_YEAST
12	145.5	5.5	1938	1	MYHD_HUMAN
13	141.5	5.4	1005	1	RA50_METJA
14	141.5	5.4	3433	1	UTRO_HUMAN
15	139.5	5.3	2663	1	CENE_HUMAN
16	138	5.2	5430	1	ACFT_HUMAN
17	137	5.2	1940	1	MYH3_HUMAN
18	137	5.2	2230	1	GOG4_HUMAN
19	135.5	5.1	1679	1	YIO9_YEAST
20	135	5.1	1060	1	EG51_XENLA
21	133.5	5.1	1957	1	YD86_SCHPO
22	132.5	5.0	852	1	RA50_THEMEA
23	131.5	5.0	1978	1	MYHB_CHICK
24	130.5	5.0	971	1	Y228_BORBU
25	130.5	5.0	1940	1	MYH3_RAT
26	130	4.9	793	1	KATA_ARATH
27	129.5	4.9	1427	1	REST_HUMAN
28	129.5	4.9	2077	1	TEGU_HSV60
29	129	4.9	539	1	MYS3_HYDAT
30	129	4.9	1939	1	MYH1_HUMAN
31	128	4.9	1744	1	TANA_XENLA
32	128	4.9	1853	1	MY5A_MOUSE
33	128	4.9	2704	1	BPAL_HUMAN

34	128	4.9	2869	1	BBP1_PLAVB
35	127.5	4.8	1722	1	MFPI_TOBAC
36	127.5	4.8	1875	1	MLPI_YEAST
37	127	4.8	1189	1	SCII_CHICK
38	126	4.8	1941	1	MYH2_HUMAN
39	125.5	4.8	571	1	YHY4_YEAST
40	125.5	4.8	978	1	RA50_AQUAE
41	125.5	4.8	1067	1	EG52_XENLA
42	125.5	4.8	1966	1	MYSB_CAEEL
43	125.5	4.8	1969	1	MYSB_CAEEL
44	125.5	4.8	2198	1	YLJ2_CAEEL
45	125.5	4.8	3210	1	CENF_HUMAN

ALIGNMENTS

RESULT	1	CLUS_HUMAN	STANDARD;	PRT;	449 AA.
ID	CLUS_HUMAN	STANDARD;	PRT;	449 AA.	
AC	P10909: P11380; P11381;				
DT	01-JUL-1989 (Rel. 11, Created)				
DT	01-JUL-1989 (Rel. 11, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Clusterin precursor (Complement-associated protein SP-40, 40)				
DE	(Complement cytotoxicity inhibitor) (CLI) (NAL and NA2) (Apolipoprotein J) (Apo-J) (TRPM-2).				
GN	CLU.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	TISSUE=Liver;				
RX	MEDLINE=89251601; PubMed=2721499;				
RA	Kirsbaum L., Sharpe J.A., Murphy B., D'Apice J.F.A., Classon B.,				
RA	Hudson P., Walker I.D.;				
RT	"Molecular cloning and characterization of the novel, human				
RT	complement-associated protein, SP-40, 40: a link between the				
RT	complement and reproductive systems.";				
RL	EMBO J. 8:711-718(1989).				
RN	[2]				
RP	SEQUENCE OF 2-449 FROM N.A.				
RX	MEDLINE=89386692; PubMed=2780565;				
RA	Tenne D.E., Tschoop J.;				
RT	"Molecular structure and functional characterization of a human				
RT	complement cytotoxicity inhibitor found in blood and seminal plasma:				
RT	identity to sulfated glycoprotein 2, a constituent of rat testis				
RT	fluid.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 86:7123-7127(1989).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=90344779; PubMed=1974459;				
RA	de Silva H.V., Harmony J.A.K., Stuart W.D., Gil C.M., Robbins J.;				
RT	"Apolipoprotein J: structure and tissue distribution.";				
RL	Biochemistry 29:5380-5389(1990).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=94237156; PubMed=8181474;				
RA	Wong P., Taillefer D., Lakin J., Pineault J., Chader G.,				
RA	Tenniswood M.;				
RT	"Molecular characterization of human TRPM-2/clusterin, a gene				
RT	associated with sperm maturation, apoptosis and neurodegeneration.";				
RL	Eur. J. Biochem. 221:917-925(1994).				
RN	[5]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=93186813; PubMed=7680346;				
RA	Pineault J.M., Tenniswood M.;				
RT	"Genomic organization and expression of the rat TRPM-2 (clusterin)				
RT	gene, a gene implicated in apoptosis.";				
RL	J. Biol. Chem. 268:5021-5031(1993).				
RN	[6]				

Q00798	plasmodium
Q9m7j4	nicotiana t
Q02455	saccharomyc
Q90988	gallus gall
Q9ukx2	homo sapien
P38871	saccharomyc
O67124	aquifex ae
Q91783	xenopus lae
P02566	caenorhabdi
P12844	caenorhabdi
P34367	caenorhabdi
P49454	homo sapien

RP SEQUENCE FROM N.A.
RA Blechschmidt K., Rosenthal A.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RX SEQUENCE OF 61-449 FROM N.A.
RA MEDLINE-92020896; PubMed-1924317;
RA Dank M., Chabot J.G., Mercier C., Benabid A.L., Chauvin C.,
RA Quirion R., Suh M.J.
RT "Human gliomas and epileptic foci express high levels of a mRNA
RT related to rat testicular sulfated glycoprotein 2, a purported marker
RT of cell death.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:8577-8581(1991).
RN [8]
RP SEQUENCE OF 140-449 FROM N.A.
RC TISSUE=Liver;
RA Glew M.D., Kirsbaum L., Bozas S.E., Walker I.D.;
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE OF 23-33 AND 228-240.
RX MEDLINE-91265608; PubMed-3154963;
RA Hochstrasser A.-C., James R.W., Martin B.M., Harrington M.,
RA Hochstrasser D.F., Pometta D., Merrill C.R.;
RT "HDL particle associated proteins in plasma and cerebrospinal fluid:
RT identification and partial sequencing";
RL Appl. Theor. Electrophor. 1:73-76(1988).
RN [10]
RP SEQUENCE OF 23-33; 229-242; 303-317 AND 397-403.
RX MEDLINE-91230083; PubMed-1903064;
RA James R.W., Hochstrasser A.-C., Borghini I., Martin B.M., Pometta D.,
RA Hochstrasser D.F.;
RT "Characterization of a human high density lipoprotein-associated
RT protein, NAI/NA2. Identity with SP-40,40, an inhibitor of complement-
RT mediated cytotoxicity.";
RL Arterioscler. Thromb. 11:645-652(1991).
RN [11]
RP SEQUENCE OF 23-52 AND 228-257.
RX MEDLINE-90354412; PubMed-2387851;
RA de Silva H., Stuart W.D., Park Y.B., Mao S.J.T., Gil C.M.,
RA Wetterau J.R., Busch S.J., Harmony J.A.K.;
RT "Purification and characterization of apolipoprotein J.";
RL J. Biol. Chem. 265:14292-14297(1990).
RN [12]
RP SEQUENCE OF 23-37 AND 228-242.
RX MEDLINE-90097955; PubMed-2601725;
RA Choi N.H., Mazda T., Tomita M.;
RT "A serum protein Sp40,40 modulates the formation of membrane attack
RT complex of complement on erythrocytes.";
RL Mol. Immunol. 26:835-840(1989).
RN [13]
RP SEQUENCE OF 23-41 AND 228-246.
RX MEDLINE-93319521; PubMed-8328966;
RA Ghisio J., Matsubara E., Koudinov A., Choi-Miura N.-H., Tomita M.,
RA Wisniewski T., Frangione B.;
RT "The cerebrospinal-fluid soluble form of Alzheimer's amyloid beta is
RT complexed to SP-40,40 (apolipoprotein J), an inhibitor of the
RT complement membrane-attack complex.";
RL Biochem. J. 293:27-30(1993).
RN [14]
RP PARTIAL SEQUENCE, DISULFIDE BONDS, AND CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE-92201397; PubMed-1551440;
RA Kirsbaum L., Bozas S.E., Walker I.D.;
RT "SP-40,40, a protein involved in the control of the complement
RT pathway, possesses a unique array of disulphide bridges.";
RL FEBS Lett. 297:70-76(1992).
RN [15]
RP DISULFIDE BONDS.
RX MEDLINE-93147076; PubMed-1491011;
RA Choi-Miura N.H., Takahashi Y., Nakano Y., Tobe T., Tomita M.;
RT "Identification of the disulfide bonds in human plasma protein
RT SP-40,40 (apolipoprotein-J).";
RL J. Biochem. 112:557-561(1992).
RN [16]
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE-97477902; PubMed-9336835;
RA Kapron J.T., Hilliard G.M., Laking J.N., Tenniswood M.P., West K.A.,
RA Carr S.A., Crabb J.W.;
RT "Identification and characterization of glycosylation sites in human
RT serum clusterin.";
RL Protein Sci. 6:2120-2133(1997).
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (APOPTOSIS).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X14723; CAA32847.1; -
DR EMBL: M25915; AAA35692.1; ALT_INIT.
DR EMBL: J02908; AAA51765.1; ALT_INIT.
DR EMBL: M74816; AAA60321.1; -
DR EMBL: M64722; AAB06508.1; -
DR EMBL: L00974; AAA60567.1; -
DR EMBL: AF311103; -; NOT_ANNOTATED_CDS.
DR EMBL: M63379; AAB06507.1; -
DR EMBL: M63376; AAB06507.1; JOINED.
DR EMBL: M63377; AAB06507.1; JOINED.
DR EMBL: M63378; AAB06507.1; JOINED.
DR EMBL: A21577; CAA01560.1; -
DR PIR: A35833; A35833.
DR PIR: A37816; A37816.
DR PIR: B37816; B37816.
DR PIR: A41386; A41386.
DR PIR: PLO135; PLO135.
DR PIR: PLO136; PLO136.
DR PIR: S04662; S04662.
DR PIR: S07433; S07433.
DR PIR: S09339; S09339.
DR PIR: S34056; S34056.
DR PIR: S43646; S43646.
DR SWISS-2DPAGE; P10909; HUMAN.
DR MIM; 185430; -
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; Cla; 1.
DR SMART; SM00030; CLD; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
KW Apoptosis; Complement pathway; Glycoprotein; Plasma; Sulfation;
KW Signal.
FT SIGNAL. 1 22
FT CHAIN 23 449
FT CHAIN 23 227
FT CHAIN 228 449
FT CHAIN 102 313
FT DISULFID 113 305
FT DISULFID 116 302
FT DISULFID 121 295
FT DISULFID 129 285
FT CARBOHYD 86 86
FT CARBOHYD 103 103
FT CARBOHYD 145 145
FT CARBOHYD 291 291
FT CARBOHYD 354 354
FT CARBOHYD 374 374
FT CARBOHYD 28 28
FT CONFLICT 47 47
FT CONFLICT 52 52
FT CONFLICT 305 305
FT CLUSTERIN.
FT BETA-CHAIN.
FT ALPHA-CHAIN.
FT INTERCHAIN.
FT INTERCHAIN.
FT INTERCHAIN.
FT INTERCHAIN.
FT INTERCHAIN.
FT N-LINKED (GLCNAC. . .).
FT N-LINKED (GLCNAC. . .).
FT N-LINKED (GLCNAC. . .).
FT N-LINKED (GLCNAC. . .).
FT N-LINKED (GLCNAC. . .).
FT D -> S (IN REF. 9 AND 10).
FT Q -> H (IN REF. 11).
FT G -> Q (IN REF. 11).
FT C -> M (IN REF. 10).

```

Query Match      16.5%; Score 434; DB 1; Length 449;
Best Local Similarity 25.8%; Pred. No. 2e-19;
Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

QY 30 MKPPLLVIVCLLWLDKSHCAPTWK-----DKTAISENLKSFSEVGEIDAEDEVKKALT 83
DB 1 MKMTLLLVFVGLL-----TWESQVLGDQTVSDNELQMSNQSKYVNIQNAV 51
QY 84 GIKOMKIMMERKEHTNLMTLKKCKREEKQALKLLNEVQHEERLCRESLADSWG 143
DB 52 GVQIKTLIEKNEERTKLNSLEAAKKKEDALNETRESETKLKLPGVCNMTMALWE 111
QY 144 ECRSCLNNCMRYT-TCQPSWSVSKNKKIERFRKIYQFLPFPHEDNEKDLPISEKLEK 202
DB 112 ECRPCLQKCMKFARVCRSGSLGVQGLEFLNQSPFFYFWNGDR-----IDSLLEN 165
QY 203 DAQLTQMEDV-----FSQLTVDVNSLNRNFVFRMQQEDQTFQSHFIS-DTDLTPEYF 257
DB 166 DROOTHMLDVMDHFSRASIIIDLEQDRF-----FTREPQDTY--HYLPFSLPHRRPHF 218
QY 258 FPAFSKEPMTKADLE-OCWDIPNFQFCNFVSIVSEVETITKMLKAIEDLPQDKAP 316
DB 219 F--FPKSRIVRSLMPFPYPLNFHAMQFPFLEMIHEA-----QQAMDIFHSPAF 267
QY 317 DHGGLISKMLPGQDRGLCGELDQNLSCRFKHEKCKQCAHLSDEC-----PDVPALHTEL 372
DB 268 QHPTEPIREGDQDRTVCREIRHNSTGCLRMKQCDCKREILSVDCSTNPNPQAKLREL 327
QY 373 DEAIRLVNNSQYQGLQWTRKHLEDTAYLVKMRGFGVWSVSELANOAPETEIFINSIQ 432
DB 328 DESLQVAERLTRKYNELLKQYKMLNTSLLEQNEQFNWVSRMLANTOGEDQYLRVT 387
QY 433 VVPRIHGNISKODETMMTDLSTLPSSNFTLKIPLESSESSNFVGVVAKALQHP-KEH 491
DB 388 TVAS-HTSDSDVSGVTEVVVKLFDSPTITVTPVEVSRNKPMTVAEKALQEQYRKH 446

RESULT 2
CLUST_COTJA STANDARD; PRT; 451 AA.
AC P14018;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Clusterin precursor (51.5 kDa protein).
GN T64.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroretina;
RX MEDLINE=89239492; PubMed=2541393;
RA Michel D., Gillet G., Volovitch M., Pessac B., Calothy G., Brun G.;
RT "Expression of a novel gene encoding a 51.5 kD precursor protein is
induced by different retroviral oncogenes in quail neuroretinal
cells.";
RT Oncogene Res. 4:127-136(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95262670; PubMed=7744033;
RA Michel D., Chatelain G., Herault Y., Brun G.;
RT "The expression of the avian clusterin gene can be driven by two
alternative promoters with distinct regulatory elements.";
RL Eur. J. Biochem. 229:215-223(1995).
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CELL DEATH.
CC -1- SUBUNIT: ANTI-PAALLEL DISULFIDE-LINKED HETERODIMER

```

```

(BY SIMILARITY)
-1- INDUCTION: BY DIFFERENT RETROVIRAL ONCOGENES.
-1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; X15825; CAA33823.1; -
CC EMBL; X80760; CAA56733.1; -
CC PIR; S07714; S07714. -
CC InterPro: IPR000753; Clusterin.
CC Pfam: PF01093; Clusterin; 1.
CC SMART; SM00035; Cla; 1.
CC SMART; SM00030; Clb; 1.
CC PROSITE; PS00492; CLUSTERIN_1; 1.
CC PROSITE; PS00493; CLUSTERIN_2; 1.
CC Glycoprotein; Signal.
CC SIGNAL 1 18 POTENTIAL.
CC CHAIN 19 451 CLUSTERIN.
CC CHAIN 19 230 BETA-CHAIN (SMALL SUBUNIT) (POTENTIAL).
CC CHAIN 231 451 ALPHA-CHAIN (LARGE SUBUNIT) (POTENTIAL).
CC DISULFID 98 314 INTERCHAIN (BY SIMILARITY).
CC DISULFID 109 306 INTERCHAIN (BY SIMILARITY).
CC DISULFID 112 303 INTERCHAIN (BY SIMILARITY).
CC DISULFID 117 296 INTERCHAIN (BY SIMILARITY).
CC DISULFID 125 286 INTERCHAIN (BY SIMILARITY).
CC CARBOHYD 99 99 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 141 141 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 355 355 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 375 375 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CONFLICT 437 437 A -> P (IN REF. 1).
CC CONFLICT 445 451 KONNTIE -> SRTTP (IN REF. 1).
CC SEQUENCE 451 AA; 51800 MW; 8CF811225C5B22F3 CRC64;

Query Match      15.9%; Score 418; DB 1; Length 451;
Best Local Similarity 25.7%; Pred. No. 1.8e-18;
Matches 118; Conservative 94; Mismatches 179; Indels 68; Gaps 15;

QY 63 LKSFSEVGEIDAEVKKALTGKIMMERKEKEHTNLMTLKKCKREEKQALKLLNE 122
DB 27 LKQLSAGSKYIDAENVNAINGVKMTLMDKTSKEHQAMLHTLETKKKEEAVKLALE 86
QY 123 VQHELEERLCRESLADSWGECRSCLENMCMRYT-TCQPSWSVSKNKKIERFRKIYQF 181
DB 87 KEKQLAKEQVNCNETMLSLWEECKPCLKHKTCMRVYKMKHSGSLGVQLEEFNRSPP 146
QY 182 LFPFHEDNEKDLPISEKLEK-----DAQLTQMEDVFSQLTVDVNSLNRNFVFRMQQ 236
DB 147 SIWNGERIDDLDRQQRERRFEDLEERFGLMEDVEDIFQDSTQLYGPAFFER---- 202
QY 237 EFDQTFQSHFISDTDLTEPY--FFPAFSKEPMTKADLEQCDWIDFN-----FFQLFC 285
DB 203 -----TPPFGGFREAF-VPPVQVRVHL-----VPRRLSRLEHFFQHPM 240
QY 286 NFSVSIYESVSETITKMLKAIEDLPQDKAPDH--GGLISKMLP-GODRGLCKEDLNLS 342
DB 241 HGFHRLFPQLEFMTQHMDG-----CHGAWHEHPLGFGFATSRNFSFTRVAVCRIRNSA 294
QY 343 RCRKFHEKCKQCAHLSDEC-----PDVPALHTELDEAIRLVNNSQYQGLQWTRKHLE 398
DB 295 GCLMRDECEKREILAVDCSQTPVQSQLREQFEDALRLAERTTRYDILLSAFQAEML 354
QY 399 DTAYLVKMRGFGVWSVSELANOAP-----ETEIFINSIQVVPRIHGNISKODETMMT 452
DB 355 NTSSLQLDLNQRQFGWVSRNLNTQGNDFGLQVTTVFVSK---TPNLEDP--SAPADTQVT- 408

```

QY 453 LSILPSSNFTLKIPLAESAESNFIYGVVAKALQHFKEH 491
 Db 409 VQLFSEPLSTVPGDISWDDPRFMEIVAEQALQHYKQ 447

RESULT 3
 CLUS_RABIT
 ID CLUS_RABIT STANDARD; PRT; 447 AA.
 AC Q9XSC5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Clusterin precursor (Apolipoprotein J) (Apo-J).
 GN CLU.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAPANESE WHITE;
 RA Miyata M., Biro S., Kaleda H., Eto H., Orihara K., Kihara T.,
 RA Obata H., Matsushita N., Matsuyama T., Tei C.;
 RT "Apolipoprotein J/clusterin is induced in vascular smooth muscle cells
 by balloon injury";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
 CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
 CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
 CC CELL DEATH (BY SIMILARITY).
 CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
 CC -!- SUBCELLULAR LOCATION: PRESENT IN CHROMAFFIN GRANULES.
 CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; AF118852; AAD24461.1; -
 DR InterPro: IPR000753; Clusterin.
 DR Pfam: PF01093; Clusterin; 1.
 DR SMART; SM00035; Cla; 1.
 DR SMART; SM00030; Clb; 1.
 DR PROSITE; PS00492; CLUSTERIN_1; 1.
 KW Glycoprotein; plasma; signal.
 FT SIGNAL 1 22
 FT CHAIN 23 447
 FT CHAIN 23 447
 FT CHAIN 225 447
 FT CHAIN 226 447
 FT DISULFID 101 311
 FT DISULFID 112 303
 FT DISULFID 115 300
 FT DISULFID 120 293
 FT DISULFID 128 283
 FT CARBOHYD 85 85
 FT CARBOHYD 102 102
 FT CARBOHYD 144 144
 FT CARBOHYD 289 289
 FT CARBOHYD 326 326
 FT CARBOHYD 352 352
 FT CARBOHYD 372 372
 SEQUENCE 447 AA; 51850 MW; COCA338EEB08590A GRC64;

Query Match 15.7%; Score 413.5; DB 1;
 Best Local Similarity 24.1%; Pred. No. 3.4e-18;
 Matches 111; Conservative 93; Mismatches 204;
 Indels 53; Gaps 11;
 QY 52 TWKDKTAISENLKSFSEVGEIDA-----DEEVKALTGKIMMERKEHTNLMST 105

Db 13 SWERGVLGDQLVSDNELQEMSTQSGKYIDREIQNAVKGVEIKTIETNEERKTLISV 72
 QY 106 LKCKREEKQEAALKLNEVQEHLEERLCRESLADSWGECRSCLENNCMRIYT-TCQPSW 164
 Db 73 LEEAKNKEDALNETRDSKTKAPPEVCNETMMALWEECKPCLKQTKMFKYARVCRSGS 132
 QY 165 SSVKNIERFFRKIYQFLPFPHEDNEKDLPISEKLEKDAQLTQMEDVFSQLTVDVNSLF 224
 Db 133 GLVGROLEEFLNQSSPFYEWINGR-----IDSLLENDROQSHVLDV-----MODSF 179
 QY 225 NBSFNVRQMQQEFDTQSFHSISDPLTEPYFFPAFSKEPMTKADLEOCWDIP----- 278
 Db 180 NRATGI---MDELFDQRFFTHKPDQTFYHSPFSY---FRPPPLHYAKSLRVNIMPISLYG 234
 QY 279 --NFFOLFCNFVSIVYESVETITMKLKAIEDLPKQD---KAPDHGGLISKMLPGDRLG 333
 Db 235 PLNFQDMFQPFEMIHQAQAMDVHLHSPAYQTPNVFEITGGPD-----DRAV 282
 QY 334 CGELQNLSCPKFHEKCKOQAHLSDC-----PDVPAHLTELDEAIRLVNVSNOQYQI 389
 Db 283 CREIRHNSGCLRMKDQCAKCEILSDVDCSANNPSONQLRQELNDSRLAEELTKRYNEL 342
 QY 390 LQWTRKHLEDATVLYEKMGRGQGWSELANQAPETETIIFNSIQVVPRIHEGNISKODETM 449
 Db 343 LQYQWKMLNTSLLDQPNQFNWVSQLANLTQGPQYLYRVSTVTS-HTSEASEPSRVT 401
 QY 450 MTDLISLPSSNFTLKIPLAESAESNFIYGVVAKALQHFKE 490
 Db 402 EVVVKLFSDSPITITIPPEVSRDNPKNFMTVAEKALQEYRK 442

RESULT 4
 CLUS_CANFA
 ID CLUS_CANFA STANDARD; PRT; 445 AA.
 AC P25473;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Clusterin precursor (Glycoprotein 80) (GP80).
 GN CLU.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hartmann K., Rauch J., Urban J., Parczyk K., Diel P., Pilarsky C.,
 RA Appel D., Haase W., Mann K., Weller A., Koch-Brandt C.;
 RT "Molecular cloning of gp 80, a glycoprotein complex secreted by
 kidney cells in vitro and in vivo. A link to the reproductive system
 and to the complement cascade";
 RL J. Biol. Chem. 266:9924-9931(1991).
 CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
 CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
 CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
 CC CELL DEATH.
 CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
 CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; M55251; AAA30846.1; -
 DR PIR; A40018; A40018.
 DR InterPro: IPR000753; Clusterin.
 DR Pfam; PF01093; Clusterin; 1.


```
DR SMART: SM00035; CLa; 1.
DR SMART: SM00030; CLb; 1.
DR PROSITE: PS00492; CLUSTERIN_1; 1.
DR PROSITE: PS00493; CLUSTERIN_2; 1.
DR Glycoprotein; Plasma; Signal.
FT SIGNAL 1 22
FT CHAIN 23 445
FT CHAIN 23 445
FT BETA-CHAIN (P35).
FT ALPHA-CHAIN (P45).
FT INTERCHAIN (BY SIMILARITY).
FT DISULFID 102 309
FT DISULFID 113 301
FT DISULFID 116 298
FT DISULFID 121 291
FT DISULFID 129 281
FT CARBOHYD 86 86
FT CARBOHYD 103 103
FT CARBOHYD 145 145
FT CARBOHYD 277 277
FT CARBOHYD 287 287
FT CARBOHYD 350 350
FT CARBOHYD 370 370
SQ SEQUENCE ~445 AA; 51789 MW; 023A37266ABEF374 CRC64;

Query Match 15.68; Score 410.5; DB 1; Length 445;
Best Local Similarity 23.9%; Pred. No. 5.1e-18;
Matches 117; Conservative 102; Mismatches 192; Indels 79; Gaps 15;

QY 30 MKPPLLVFVCLMLKDKSHCAPTKWKTAISENLKSFSEVGEIDAD-----EEVKKALT 83
Db 1 MKKTLVLVGLLL-----TWNGRVLDGQAVSDTELOEMSTEGSKYINKEIKALK 51
QY 84 GIKOMKIMMERKEHTNMLSTLKKREKQEAOLKLLNEVQHEEERLCRESLADSWG 143
Db 52 GVGIKTLIEOTNEERKSLNLEEAKKKEDALNDTKDSETKLKASQGVNDFMALWE 111
QY 144 ECRSCLENNCRMYT-TCQPSWSVKNKIERPFKIVQFLPPPHEDNEKDLPISEKLIK 202
Db 112 ECKPCLAKOTCMKFYARVCRSGSLVGHQLEEFLLNQSSPFYFWMGDR-----IDSLLEN 165
QY 203 DAQLTOMEDVFSOLTVDVNSLFNRSVNFVROMQOEFDOTFQSHFIS----DITDLEPYFF 258
Db 166 DROOTHALDV-----MQDSFNRASSI-----MDELQDFRFTREPQDTHYSP--F 209
QY 259 PAFSKEPMTKADLEQCWDI-----PNFQLFCNFSVSIYESVSTITMKLKAIB-- 307
Db 210 SLFQRPFPPFKFIARNIIPFRFQPLNFHDMFQFPFDMHQ-AQAMDVNLHRIPIYE 268
QY 308 --DLPKODKAPDGHGLISKMLPGODRCLGCLDONLSRCFKFHEKCKQCAHLESDC--- 362
Db 269 PIEFPEED-----NRTVCKEIRHNSTGCLMKDKQCEKQOEILSVDCSSN 312
QY 363 -PDVPALHTELDAIRLVNVSNOQYGQILQWTRKHLHEDTAYLVKEMRGQFGWSELANQA 421
Db 313 NPAQVQLRQELNSLQIAEFTKLYDELQSYQEKMENTSLKQLNEQSWSQLANLT 372
QY 422 PETEIIINSQVY-PRTHEGNISKQDQDTMTDLSILPSSNFTLKIPLEESAESSNFTGYV 480
Db 373 QSEDPFYLVQTVGTSQSDSNVPVGFVKV--VKLFSDSDPITVWIPVSRNPNKPMETV 430
QY 481 VAKALOHEKE 490
Db 431 AERALKQEIYRQ 440

RESULT 5
CLUS_MOUSE
ID CLUS_MOUSE STANDARD; PRT; 448 AA.
AC Q06890;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Clusterin precursor (Sulfated glycoprotein 2) (SGP-2) (Clusterin)
```

```
DE (Apolipoprotein J) (Apo-J).
GN CLU OR MSGP-2 OR APOJ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=93356785; PubMed=8352774;
RA Lee K.-H., Ji Y.-M., Lim H.M., Lee S.-C., You K.-H.;
RT "Molecular cloning and sequencing of sulfated glycoprotein-2 cDNA
RT from testis of mouse: implications of two different mRNAs of SGP-2.";
RL Biochem. Biophys. Res. Commun. 194:1175-1180(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93359508; PubMed=8354695;
RA French L.E., Chonn A., Ducrest D., Baumann B., Belin D., Wohlwend A.,
RA Kiss J.Z., Sappino A.P., Tschopp J., Schifferli J.A.;
RT "Murine clusterin: molecular cloning and mRNA localization of a gene
RT associated with epithelial differentiation processes during
RT embryogenesis.";
RL J. Cell Biol. 122:1119-1130(1993).
RN [3]
RP SEQUENCE FROM N.A.
RA Hodgdon B.A., Min B.H., Yan H., Farris J.A., Foster D.N.,
RA Strauch A.R.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CHA; TISSUE=Liver;
RX MEDLINE=94223204; PubMed=8169523;
RA Jordan-Starck T.C., Lund S.D., Witte D.P., Atonow B.J., Ley C.A.,
RA Stuart W.D., Swertfeger D.K., Clayton L.R., Sells S.F., Paigen B.;
RT "Mouse apolipoprotein J: characterization of a gene implicated in
RT atherosclerosis.";
RL J. Lipid Res. 35:194-210(1994).
CC -1- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A
CC VARIETY OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS,
CC MEMBRANES, AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED
CC WITH PROGRAMMED CELL DEATH.
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN STOMACH, LIVER, BRAIN, AND
CC TESTIS, WITH INTERMEDIATE LEVELS IN HEART, OVARY, AND KIDNEY.
CC -1- PTM: EXTENSIVELY GLYCOSYLATED WITH SULFATED N-LINKED
CC CARBOHYDRATES.
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D14077; BAA03162.1; -
DR EMBL; L08235; AAA37422.1; -
DR EMBL; L05670; AAA37284.1; -
DR EMBL; S70244; AAB30623.1; -
DR PIR; A40714; A40714.
DR PIR; JN0699; JN0699.
DR MGD; MGI:88423; Clu.
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; CLa; 1.
DR SMART; SM00030; CLb; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
DR Glycoprotein; Sulfation; Signal.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 448 CLUSTERIN.
FT CHAIN 22 226 BETA-CHAIN (SMALL SUBUNIT)
```


Db	233	----	NFHDVQFP-----	YDMIHOQAQMDAHLQ-----	RTPVHFFPTM-EFTENDRWTC	277
Qy	335	GELDNLSRCFKFEHKCQKQAHLSDC-----	PDVPALHTELDEARLVNVSNOOYQGL	390		
Db	278	KEIRHNSGTCLRMKDQCEKQCEILEYDCSASNPQTLLRQQLNASLQAEKFSRLYDQLL	337			
Qy	391	QMTRRKHLLEDYAILVEKMRGQFGWVSELANOAPTE-----	IFNSYQVVPRIHEGNTSK	444		
Db	338	QSYQOKMLNLSALLKQLNEQFTWVSQLANLTQSDDDHYLQVETVNSHSDPSIPSG-----	393			
Qy	445	QDETHMTDLILPSSNFTLKIPLEESAESNSFTGYVAKALQHKF	489			
Db	394	-----	LTKVTVKUFNSFPFTVTPQEVSSPNFMENYAEKALQOYR	433		

```

RESULT 7
CLUS_PIG
ID CLUS_PIG STANDARD; PRT; 446 AA.
Q29549;
DT 15-JUL-1998 (Rel. 36, Created)
DT 13-JUL-1998 (Rel. 36, Last sequence update)
DT 13-JUL-1998 (Rel. 36, Last annotation update)
DE Clusterin precursor (Complement cytotoxicity inhibitor) (CLI).
GN CLU.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
[1]
SEQUENCE FROM N.A.
RN RP MEDLINE=92184774; PubMed=1544909;
RA Diemer V., Hoyle M., Baglioni C., Millis A.J.;
RT "Expression of porcine complement cytotoxicity inhibitor mRNA in
RT cultured aortic smooth muscle cells. Changes during differentiation
RT in vitro."
RL J. Biol. Chem. 267:5257-5264 (1992).
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN BRAIN AND LIVER. LOWER
CC LEVELS ARE DETECTED IN OTHER TISSUES, INCLUDING THE AORTA.
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M84639; AAA31013.1; -.
CC InterPro: IPR000753; Clusterin.
CC Pfam: PF01093; Clusterin; 1.
CC SMART: SM00035; CLa; 1.
CC SMART: SM00030; CLD; 1.
CC PROSITE: PS00492; CLUSTERIN_1; 1.
CC PROSITE: PS00493; CLUSTERIN_2; 1.
CC Glycoprotein; Signal.
FT SIGNAL 1 28 BY SIMILARITY.
FT CHAIN 29 446 CLUSTERIN.
FT CHAIN 29 227 BETA-CHAIN (A CHAIN).
FT CHAIN 228 446 ALPHA-CHAIN (B CHAIN).
FT FT DISULFID 102 312 INTERCHAIN (BY SIMILARITY).
FT FT DISULFID 113 304 INTERCHAIN (BY SIMILARITY).
FT FT DISULFID 116 301 INTERCHAIN (BY SIMILARITY).
FT FT DISULFID 121 294 INTERCHAIN (BY SIMILARITY).
FT FT DISULFID 129 284 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT	CARBOHYD	145	145	N-LINKED (GLCNAC..)	(POTENTIAL).
FT	CARBOHYD	290	290	N-LINKED (GLCNAC..)	(POTENTIAL).
FT	CARBOHYD	316	316	N-LINKED (GLCNAC..)	(POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAC..)	(POTENTIAL).
FT	CARBOHYD.	373	373	N-LINKED (GLCNAC..)	(POTENTIAL).
SQ	SEQUENCE	446 AA;	51774 MW; B1DSB434B668E3AA	CRC64;	
 Query Match Best Local Similarity 15.2%; Score 400.5; DB 1; Length 446; Matches 118; Conservative 98; Mismatches 196; Indels 71; Gaps					
Qy	34	LLVIVICLLMLKDHCAPTWK-----DKTAISENLKSFSEVGIEDADDEVKKALTGK 86	: : : : : :	:	:
Dd	4	LLLLVLGLLL-----TWNGPWVLGDKAISDELQEMSTEGSKYVNKEINAKLEVK 54	: : : :	:	:
Qy	87	OMKTMMERKEKEHNLMSTLTKCKREEKOEAALKLNNEVOEHLEEFEEELCRSLADSMGEGR 146	: : : : : :	:	:
Dd	55	QIKTLIQSNEERKSLLSSLEAEAKKKEDALNDTRDTETKLKGSGQLCNETMMAWEEK 114	: : : :	:	:
Qy	147	SCLENNCMRIYT-TCQPSSWSGVKNKIEFRPKRYQFLFPHEDNEXKDLPISKLIKEDAQ 205	: : : : : :	:	:
Dd	115	PCLAQTCMKFYAVCRSGSLGVGHOLEEFNLQSPPFWINGDR-----IDSLMENDRQ 168	: : : :	:	:
Qy	206	LTMEDVFSQLTVDVNSLFNRSFNVFMQQOEFDTFQSHFIS----DTDLTEPY-----256	: : : : : :	:	:
Dd	169	QSHVMDI-----MEDSFNRASNI-----MDLEFDQRRFNREFDTPFFSPFGSSH 214	: : : :	:	:
Qy	257	--FPFAPSKPEMTKADLECWDIPNFQLFCNFSYSIESVSETITKMLKAIEDLPKQD 313	: : : : : :	:	:
Dd	215	GSLGFNPFKSRFA RNIMPFPLTDL-NYHDMFQPF---FDMIHQAOAMD AHLHRIPYH -268	: : : :	:	:
Qy	314	KAPDHGGILSKMLP--QODRGLCGELDONLSRCEKFPEHKCQQAHLSDPC----PDVPA 367	: : : : : :	:	:
Dd	269	-FPAG-----VPENSDRAVCKEIRHNSTGCRLMKDCQCEKREILSVDCSNASMSQM 321	: : : :	:	:
Qy	368	LHTDELAIRLVNVSNOQYGQIQLMQTKHLEDATAYLEVKMARGOGGWSELANOAPETEII 427	: : : : : :	:	:
Dd	322	LRQEITYTSLOMAEKFSKLYLDQLSQYOQKMLNTSSLKQLNQEQFSWVSOIANLTQNDDRY 381	: : : :	:	:
Qy	428	FNSTQVYPRIHEGNISKQDETMMTDLSILPSSNFTLKIPLEESAESSNFTGYVVAKALQH 487	: : : : : :	:	:
Dd	382	YLQVTTV-NSHGSDPSVPESGLTKVVKVLFDPSPITLIPOEV-S-DPKFMETVAEEALQQ 438	: : : :	:	:
Qy	488	FKE 490	:	:	:
Dd	439	YRQ 441	:	:	:
 RESULT 8 CLUS_RAT . STANDARD; PRT; 447 AA. ID CLUS_RAT .					
AC	P05371;				
DT	01-NOV-1988 (Rel. 09, Created)				
DT	01-FEB-1994 (Rel. 28, Last sequence update)				
DT	01-OCT-1996 (Rel. 34, Last annotation update)				
DE	Clusterin precursor (Sulfated glycoprotein 2) (SGP-2) (Dimeric acid glycoproteins) (DAG) (Testosterone repressed prostate message-2) (TRPM 2).				
GS	Clu.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
[1]					
RN	SEQUENCE FROM N.A., AND SEQUENCE OF 22-45 AND 227-241.				
RX	MEDLINE=88000523; PubMed=3651384;				
RA	Collard M.W., Griswold M.D.				
RT	"biosynthesis and molecular cloning of sulfated glycoprotein 2 secreted by rat Sertoli cells.";				
RL	Biochemistry 26:3297-3303(1987).				
[2]					
RN	SEQUENCE FROM N.A.				

RA Pineault J.M., Tenniswood M.;
RN Submitted (XXX-1991) to the EMBL/Genbank/DBJ databases.
RN [3]
RX MEDLINE-89149740; PubMed-2920020;
RA Betuzzi S., Hilpakka R.A., Gilna P., Liao S.;
RT "Identification of an androgen-regressed mRNA in rat ventral prostate
RT as coding for sulphated glycoprotein 2 by cDNA cloning and sequence
RT analysis";
RL Biochem. J. 257:293-296(1989).
RN [4]
RX MEDLINE-88326333; PubMed-3415696;
RA Cheng C.Y., Chen C.C., Feng Z., Marshall A., Bardin C.W.;
RT "Rat clusterin isolated from primary Sertoli cell-enriched culture
RT medium is sulfated glycoprotein-2 (SGP-2).";
RL Biochem. Biophys. Res. Commun. 155:398-404(1988).
RN [5]
RX CHARACTERIZATION OF TRPM-2.
RX MEDLINE-90134121; PubMed-2299741;
RA Bandyk M.G., Sawczuk I.S., Olsson C.A., Katz A.E., Buttyan R.;
RT "Characterization of the products of a gene expressed during
RT androgen-programmed cell death and their potential use as a marker
RT of urogenital injury";
RL J. Urol. 143:407-413(1990).
CC -!- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH.
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED BY CELLS UNDERGOING PROGRAMMED
CC DEATH AS A RESULT OF THE HORMONAL STIMULI OR A TRAUMATIC INSULT.
CC -!- PTM: EXTENSIVELY GLYCOSYLATED WITH SULFATED N-LINKED
CC CARBOHYDRATES.
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M16975; AAA41273.1; -
CC EMBL: M64723; AAA42298.1; -
CC EMBL: M64733; AAA42299.1; -
CC EMBL: X13231; CAA31618.1; -
CC PIR: A27205; A27205.
CC PIR: A31575; A31575.
CC PIR: B31575; B31575.
CC PIR: A45890; A45890.
CC PIR: S18491; S18491.
CC InterPro: IPR000753; Clusterin.
CC Pfam: PF01093; Clusterin; 1.
CC SMART: SM00035; Cla; 1.
CC SMART: SM00030; Clb; 1.
CC PROSITE: PS00492; CLUSTERIN_1; 1.
CC PROSITE: PS00493; CLUSTERIN_2; 1.
KW Sulfation; Glycoprotein; Spermatogenesis; Signal.
FT SIGNAL 1 21
FT CHAIN 22 447 CLUSTERIN.
FT CHAIN 22 447 BETA-CHAIN (SMALL SUBUNIT).
FT CHAIN 227 447 ALPHA-CHAIN (LARGE SUBUNIT).
FT DISULFID 101 312 INTERCHAIN (BY SIMILARITY).
FT DISULFID 112 304 INTERCHAIN (BY SIMILARITY).
FT DISULFID 115 301 INTERCHAIN (BY SIMILARITY).
FT DISULFID 120 284 INTERCHAIN (BY SIMILARITY).
FT DISULFID 128 284 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (PROBABLE).

FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CONFLICT 187 187 D -> H (IN REF. 1).
SQ SEQUENCE 447 AA; 51375 MW; 9E2FA33E5E0C146E CRC64;

Query Match 14.4%; Score 380; DB 1; Length 447;
Best Local Similarity 23.5%; Pred. No. 3.6e-16;
Matches 113; Conservative 103; Mismatches 197; Indels 68; Gaps 15;

QY 34 LLVIVCLLWLDKSDCAPTWKDKTAISENLKSFSEGEID-----ADEVKKALTGKIQ 87
Db 4 LLLCVALL-----TWDNCGWVLGEQEFSDNELQELSTQGSRYVNEIQNAVGVKH 54
QY 88 MKIMMERKEKEHTLMSTLTKCKREEKQEAALKLLNEVQEHLEEEERLCRLSADSWGECS 147
Db 55 IKTLEIKTNAERKSLNLEAEAKKKKEGALDDTRDSEMKLKAPPEVCNETMMALWEECKP 114
QY 148 CLENNCMRIYT-TCQPSWSSVKNKIERFRKIYQFLFFPHEDNEKDLPISEKLIKDAQL 206
Db 115 CLKHTCMKFYARVCRSGSLVGRQLLEFLNQSSPFYFMNGDR-----IDSLLESDRQQ 168
QY 207 TQ----MEDVFSOLTVDVNSLFRNSFNVRQMQEFDQTFQSHFISDTDLTEPYFPP--- 259
Db 169 SQVLDAMQDSFTRASCIIDTLFQDRF--ETHEPQDI-----HHF-----SPMGFPK 214
QY 260 ---AFSKEPMTKADLEQCWDIP-NFFQLCNFSVSIVSEYSETITKMLKAIEDLPKDKA 315
Db 215 PHFLYPKSLVRLSPLSHYGLPSLPHNMPFPFMDIHOQAQAMDVLHSPALQFPDVFDEL 274
QY 316 PDHGLISLMLPGQ-DRGLCGELDQNLSCFPHEKCKCOAHLSEDC-----PDVPAHHT 370
Db 275 KE-----GEDDPTVCYKEIRHNSSTGLKMKGCQCEILSVDCSTNNPAQANLRQ 324
QY 371 ELDEAIRLVNSNOQYGOILQMTKRLHLEDTAYLVKMRQGFQGWVSELN--QAPETEILF 428
Db 325 ELNDSLOVAERLTQQYNELLHLSLOKMLNTSSLLEQLNDQFTWVSQLANTQGDQYLRV 384
QY 429 NSIQVVRPIHESNISKQDETMMDTSLPSSNFTLKIPLEESAESSNFTGYGVAKALQHF 488
Db 385 STVTT-----HSSDSEVPSPRVTEVVVKLFDSPTITVVLPPEVSKDNPKFMDTVAEKALQ 440
QY 489 K 489
Db 441 R 441

RESULT 9
CLUS_MESAU
ID CLUS_MESAU STANDARD; PRT; 191 AA.
AC P14883;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Clusterin (Sulfated glycoprotein 2) (SGP-2) (Fragment).
GN CLU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE OF 24-123 FROM N.A.
RX MEDLINE-89386721; PubMed-2780570;
RA Duguid J.R., Bohmont C., Liu N., Tourtellotte W.W.;
RT "Changes in brain gene expression shared by scrapie and Alzheimer
RT disease";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7260-7264(1989).
RN [2]
RP SEQUENCE OF 1-23 AND 124-191 FROM N.A.
RA Duguid J.R., Bohmont C., Liu N., Tourtellotte W.W.;
RL Submitted (AUG-1989) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY


```

DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR Pfam: PF01576; Myosin_tail; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS00096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW Multigene family.
FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
FT DOMAIN 785 814 IQ.
FT DOMAIN 843 1938 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 659 681 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 761 775 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 130 130 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 699 699 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 709 709 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1938 AA; 223678 MW; 1F6D006416381CD5 CRC64;

Query Match 5.5%; Score 145.5; DB 1; Length 1938;
Best Local Similarity 20.7%; Pred. No. 0.27;
Matches 100; Conservative 93; Mismatches 206; Indels 83; Gaps 17;

QY 59 ISENKSFSEVGEIDAEDEV-KVLTGKIMMERKEHEHNLMSTLKKCR-----EEK 113
DB 1150 ISELEAS--GATSAQIENKREAFKMRRLDEATLQHEATATLRKKQADSVAEL 1207
QY 114 QEALKLLNEQVHELEERERLCRESLAD--SWGECRSLENNCHRIYTCQPSSSVKNKI 171
DB 1208 GEQIDNQRVKQLEKESEKLEIDMNASNIEALSXSNIERTCTVQEDQSEIKAKD 1267
QY 172 ERFRRKIYQFLFPFHEDNEKDLPISEKLIIEKDAQLTQMEDVFSQLTVDVNSLFNRSPNVF 231
DB 1268 EQQTOLI-----HDLNMQKARLTQNGELSHRVEKESLSQLTKSKQALTOQLEELK 1320
QY 232 RQMQQEFD-QTFQSHFTS-----DTDLTPEYFFAFKSEPKMTKADLEOCW-----DIPNFF 281
DB 1321 ROMEETKAKNAHAHQAQSSRHPCDLLR-----EQYEEQEAQAEQLQALSKANSEVAQWK 1376
QY 282 QLCNFSVSIVSEYSETITKMLKAIEDL-PKQDKAPDHGLISK---MLPGQDGRGLGEL 337
DB 1377 TKYETDAIQTEELEEAQKLLAQLQAEAEKTEFTANSKASLEKTKQRLQGEVEDLMROL 1436
QY 338 DONLSRC-----FKFHEKQKQAHLSDCPDVPAHTELDPAIRLVNVS 383
DB 1437 ERSHTACATLDKKORNFQKVLAEWKQKLDSEQAEEAAQKESRSLSTEL---FKMRNAYE 1493
QY 384 QYQYQI--LQWTKHLEDYALVEKMRQGWSELANQAPET-----EII 427
DB 1494 EVDQETLRENNKIQEE-----ISDLTEQIAETGKNIQAEKTKKLVQEE 1540
QY 428 FNSIQVPRTHEGNISKQDETMM---TDLISLPSSNFTLKIPLSEASNESSNFIYGVVAKA 484
DB 1541 KSDQVALEVEGSLSEHSKILRVQLELSQVKSSELDKRVKIEKDETEQLKRNSQRAAEA 1600
QY 485 LQ 486
DB 1601 LQ 1602

RESULT 13
RA50_METJA STANDARD; PRT; 1005 AA.
AC Q58718;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR MJ1322.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL jannaschii."
CC Science 273:1058-1073(1996).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mrell complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mrell by unwinding
CC and/or repositioning DNA ends into the mrell active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mrell (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: U67572; AAB99331.1; -
CC TIGR: MJ1322; -
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR001687; ATP_GTP_A.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02483; SMC_N; 1.
DR Pfam: PF02463; SMC_N; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 32 39 ATP (BY SIMILARITY).
FT DOMAIN 158 849 COILED COIL (POTENTIAL).
SQ SEQUENCE 1005 AA; 119387 MW; 9BBB848173E788F3 CRC64;

Query Match 5.4%; Score 141.5; DB 1; Length 1005;
Best Local Similarity 20.0%; Pred. No. 0.22;
Matches 81; Conservative 70; Mismatches 131; Indels 123; Gaps 16;

QY 54 KDKTAISENL-----KSFSEVGEIDAEDEVKALTGKIMMERKEHEHNLMST 105
DB 419 QEKSKTEKNINDLETRINKLLEETKNIDI-ESTENSKSEIEKKVLENQKEIKLNKK 477
QY 106 LKCKREEQKQALKLLNEVQEHLEEEERLCRESLADSGECRSCL-----ENNCRIYTCQ 161
DB 478 LGEINSE---IKRLKKILDDELKEVE-----GKPLCKPTIDENKKMELINQHK 522
QY 162 PSWSSVKNKIERFERFKIYQFLFPFHEDNEKDLPISEKLIIEKDAQLTQMEDVFSQLTVDVN 221
DB 523 TQLNNKYTELEIEINKKIREI-----EKDIEKLEKIDKEENLTKLYLEKQSQIE 574
QY 222 SLFNRSENVFQMQQEDQFQSHFISDTDLTETPPFPFAPFSKEPMTKADLEQCWDIPNFF 281
DB 575 ELEKLNK-YKEQLDEINKKISNVVINGKPVDE-----ILE---DIKSQL 615
QY 282 QLCFNF-----SVSIYESVSET-ITKMLKAIEDLPKQDKAPDHGGLISKMLPGQDGRGLC 334

```


Db 616 NFKFNYNOYLSAVSLNSVDEGINRRIKTEINI-----VSGWNK----- 656
QY 335 GELDQNLSCFRKHEKQKQAHLSDCDPVPALHTELDE-----AIRLVNSNQ----- 384
Db 657 -----EKCREELNKLREDEIRNRLKDKLNELKKEKELIEIENRRSLKFD 702
QY 385 QYQOILQMTKRKULEDTAYLVKMRGQFGWSELANOAPETEIEFN 429
Db 703 KYKEYLGLTEK-----LEELKNIKDGLLEIYN 729

RESULT 14
UTRO_HUMAN
ID UTRO_HUMAN STANDARD; PRT; 3433 AA.
AC P46939;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Utrrophin (dystrophin-related protein 1) (DRP1) (DRP).
GN UTRN OR DMDL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93096045; PubMed=1461283;
RA Tinsley J.M., Blake D.J., Roche A., Fairbrother U., Riss J.,
RA Byth B.C., Knight A.E., Kendrick-Jones J., Suthers G.K., Love D.R.,
RA Edwards Y.H., Davies K.E.;
RT "Primary structure of dystrophin-related protein.";
RL Nature 360:591-593(1992).
RN [2]
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 144-261.
RX MEDLINE=99141377; PubMed=9887274;
RA Keep N.H., Norwood F.L.M., Moores C.A., Winder S.J.,
RA Kendrick-Jones J.;
RT "The 2.0-A structure of the second calponin homology domain from the
actin-binding region of the dystrophin homologue utrophin.";
RL J. Mol. Biol. 285:1257-1264(1999).
CC -!- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE
PLASMA MEMBRANE (BY SIMILARITY TO DYSTROPHIN).
CC -!- SUBCELLULAR LOCATION: NEUROMUSCULAR JUNCTION.
CC -!- TISSUE SPECIFICITY: MUSCLE.
CC -!- SIMILARITY: STRONG, TO DYSTROPHIN.
CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
ABP-120, ABP-180, OR BETA-FODRIN).
CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 WW DOMAIN.
CC -!- SIMILARITY: CONTAINS 20 SPECTRIN REPEATS.
CC -!- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; X69086; CA448829.1; -
DR PIR; S28381; S28381.
DR PDB; 1BHD; 16-FEB-99.
DR MIM; 128240; -
DR InterPro; IPR001589; Actinin_act_bind.
DR InterPro; IPR001715; Calponin_hom.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR003122; TarH.
DR InterPro; IPR001202; WW.
DR InterPro; IPR002349; WW_domain.
DR InterPro; IPR000433; ZnF_Z2.

DR Pfam; PF00307; CH; 2.
DR Pfam; PF00435; Spectrin; 19.
DR Pfam; PF00397; WW; 1.
DR Pfam; PF00569; Z2; 1.
DR PRINTS; PR00403; WWDOMAIN.
DR SMART; SM00033; CH; 2.
DR SMART; SM00150; SPEC; 18.
DR SMART; SM00319; TarH; 1.
DR SMART; SM00456; WW; 1.
DR SMART; SM00291; ZnF_Z2; 1.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00020; ACTININ_2; 1.
DR PROSITE; PS00021; CH; 2.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PROSITE; PS00020; WW_DOMAIN_2; 1.
DR PROSITE; PS01357; ZF_Z2_1; 2.
DR PROSITE; PS01357; ZF_Z2_2; 1.
KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
KW Repeat; 3D-structure; Zinc-finger.
FT DOMAIN 1 246
FT ACTIN-BINDING.
FT DOMAIN 31 135
FT CH 1.
FT CH 2.
FT REPEAT 150 252
FT SPECTRIN 1.
FT REPEAT 253 308
FT SPECTRIN 2.
FT REPEAT 309 417
FT SPECTRIN 3.
FT REPEAT 418 526
FT SPECTRIN 4.
FT REPEAT 541 637
FT SPECTRIN 5.
FT REPEAT 687 798
FT SPECTRIN 6.
FT REPEAT 803 902
FT SPECTRIN 7.
FT REPEAT 1016 1083
FT SPECTRIN 8.
FT REPEAT 1125 1230
FT SPECTRIN 9.
FT REPEAT 1248 1334
FT SPECTRIN 10.
FT REPEAT 1432 1541
FT SPECTRIN 11.
FT REPEAT 1544 1649
FT SPECTRIN 12.
FT REPEAT 1652 1753
FT SPECTRIN 13.
FT REPEAT 1910 1968
FT SPECTRIN 14.
FT REPEAT 1976 2081
FT SPECTRIN 15.
FT REPEAT 2258 2333
FT REPEAT 2399 2440
FT SPECTRIN 16.
FT REPEAT 2443 2556
FT SPECTRIN 17.
FT REPEAT 2559 2636
FT SPECTRIN 18.
FT REPEAT 2658 2688
FT SPECTRIN 19.
FT REPEAT 2691 2797
FT SPECTRIN 20.
FT WW.
FT DOMAIN 2812 2845
FT ZN_FING 3064 3111
FT Z2-TYPE.
SQ SEQUENCE 3433 AA; 394488 MW; EAEBDBA09F8585B CRC64;

Query Match 5.4%; Score 141.5; DB 1; Length 3433;
Best Local Similarity 22.2%; Pred. No. 0.9; Mismatches 176; Indels 83; Gaps 18;
Matches 94; Conservative 71

QY 45 KDSHCAPTWKDKTAISENLKSFSEVGEIDAEVKKALTIGIKQKIMMERKEKHTNLS 104
Db 1551 KEAASLSEWLSATATELVOKSTSEGLGLDLDTEISWAKNVKLDLE-----KRKADLNTIT 1605
QY 105 TLKKREKQKALKLLNEVQHEERLCRESLADSGECSCLENENCMRYITTCQPSW 164
Db 1606 -----ESSAALQNLIEGSEPI-LEERLC--VLNAGWSRVRTWTDWCNTL----- 1647
QY 165 SSVKNKIERFFRKI-----YQFLFPFHEDNEKDLPISEKLIEX-----DAQLTQMED 211
Db 1648 MHQNOLEIFDGNVAHISTWLYQAEALDDEIEKKPTSQEEIVKRLVSELDANL-QVEN 1706
QY 212 VFSQLTVDVNSLNSFNVMFROMQOEFQDTFO--SHFISDTDLTEPFYFPFAFSKEPMTKA 269
Db 1707 VRDQALILMNARGSSSRELVEPKLAELNRNFEKVSQHIKSAKL-----LIAQEP---- 1755
QY 270 DLEQWDIPNFFQLFCNFSVSVIYESVETITMKLAI-----BDLPKQKAPDHGGLSKM 325
Db 1756 -LYOCLVTTTETFTGTGVPF--SDLEKLENDIENMLKFEVKHLESDDEKMDSESAQIEEV 1812
QY 326 LPQDGRGLCGELDQNLSCFCFKFHEKCKQCOAHLSDCDPVPALHTELDEAIRLVNSNQ 385

Db 1813 LQGEEMLHQPMDN-----KKEIKRLQLL-----LHTRYNK-IRKAIQQRK 1855
 QY 386 YGQILQWTRKHLEDYAVLVKMRGQFGWSELANQAPETEIEFNSIQVVPRIHEGNSIKQ 445
 Db 1856 MGQLASGIRSLPTDYLVE-----INKILLCHDDVLSLNVPELNTAIVE-DFSQ 1906
 QY 446 DETM 449
 Db 1907 EDSL 1910

RESULT 15
 CENE_HUMAN STANDARD; PRT: 2663 AA.
 AC Q02224;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Centromeric protein E (CENP-E protein).
 GN CENPE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93024922; PubMed=1406971;
 RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
 RT "CENP-E is a putative kinetochore motor that accumulates just before
 mitosis.";
 RL Nature 359:536-539(1992).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=95196755; PubMed=7889940;
 RA Thorer D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
 RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
 microtubule motor.";
 RL EMBO J. 14:918-926(1995).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.;
 RT "Characterization of the kinetochore binding domain of CENP-E reveals
 interactions with the kinetochore proteins CENP-F and HUBB1.";
 RL J. Cell Biol. 143:49-63(1998).
 CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
 KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
 OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
 AND/OR SPINDLE ELONGATION.
 CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
 CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
 CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
 QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z15005; CAA78727.1; -;
 CC PIR: S28261; S28261.
 CC HSP: P17119; 3KAR.
 CC MIM: 117143; -;
 CC InterPro: IPR001752; kinesin.
 CC Pfam: PF00225; kinesin; 1.
 CC PRINTS: PR00380; KINESINHEAVY.
 CC SMART: SM00129; KISC; 1.
 CC PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 CC PROSITE: PS00067; KINESIN_MOTOR_DOMAIN2; 1.

KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
 KW Cell cycle; Centromere.
 FT DOMAIN 1 335 KINESIN-MOTOR.
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 336 2471
 FT GLOBULAR (POTENTIAL).
 FT NP_BIND 2472 2663
 FT ATP (BY SIMILARITY).
 FT NP_BIND 86 93
 SQ SEQUENCE 2663 AA; 312087 MW; CEFCL13880C8CB8 CRC64;
 Query Match 5.3%; Score 139.5; DB 1; Length 2663;
 Best Local Similarity 18.1%; Pred No. 0.89;
 Matches 100; Conservative 87; Mismatches 175; Indels 191; Gaps 23;
 QY 59 ISENLSFSEVGGEIDAEVVKALTGK-----QMKIMMERKEKEHTNLMSTLK----- 107
 Db 1461 LKENIKEI-VAKHLETEELKVAHCLKEQETINELRVNLSKEKTEISTIOKLEAIND 1519
 QY 108 ----KCRE--EKQAL--KLLNEVQEHLEE-----PERLCRESLADSWGECRSLNNCM 154
 Db 1520 KLQNKIQEIEKEEQNLNIQISEVQENVNELKQFHKRAKDALQS-----IESKML 1572
 QY 155 RIVTTCQPSWSSVK-----NKKIERFFPKIYQFL- 182
 Db 1573 ELTNRLQESQEEIQIMKEEMKRVQEQALQIERDQKENTKEIVAKMKESQEKYQFLK 1632
 QY 183 -FPFHEDNEKDLPISSEKLIKDAQLTQMEDVFSQLTVDVNSLFNSFNVRQMQEFDQT 241
 Db 1633 MTAVNETQEKMCIEHLKEQFQKLNLENIETE-NIRLTQILHENLEEMRSVTKERD-- 1689
 QY 242 FQSHFISDITDTPYFPFAPAFSKPMYKADLEQWCQDIPNFQFCNFSVSIYESVSETIYK 301
 Db 1690 -----DLR-----SVEETLKVVERDQ-----LKENLRETIIR 1715
 QY 302 MKAIEDLPKODK-----APDHGGLISKMLPGQDRGLGELDQNLNLSRC----- 344
 Db 1716 -----DLEKQEEELKIVHMLKHEQETIDKL-----RGIVSEKTNELSNMOKLEHSNDA 1764
 QY 345 -----FKFHEKCKQCAHLSDECQDVPAL-----HTELDEAIRL 378
 Db 1765 LKAQDLKIQEELRIAHMLKEQETIDKLGRIVSEKTDKLSNMOKLENSNAKLOEKIQE 1824
 QY 379 VVNSNOQYGOILQWTRKHLEDYAVLVKMRGQFGWSELANQAPETEIEFNSIQV----- 433
 Db 1825 LKANEHQ-----LITLKKDVNETQKVVSEM-----EQKKQIKDOOSLTLSKLEIENLNL 1873
 QY 434 VPIHEG-----NISKODETMTDLSILPSSNFTLKIPLESAESSNFYGVV 481
 Db 1874 AQELHENLEEMKSVMKERDNLRRVEETIKLERD-----QLKESLOETKARDLEIQOEL 1926
 QY 482 AKALQHFKEHFT 494
 Db 1927 KTARMLSKHEKET 1939

Search completed: July 2, 2002, 11:58:14
 Job time: 367 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:48:17 ; Search time 33.04 Seconds
(without alignments)
1439,595 Million cell updates/sec

Title: US-09-722-544A-2MOD
Perfect score: 2632
Sequence: 1 MKIKAEKNEGSPSRWQLHW.....FIGYVAKALQHFKHEFTW 495
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	434	16.5	449	1 A41386	clusterin precursor
2	418	15.9	451	2 I50131	clusterin - quail
3	410.5	15.6	445	2 A40018	clusterin precursor
4	408	15.5	449	2 S07714	T64 protein precursor
5	403	15.3	448	2 A40714	clusterin precursor
6	402	15.3	448	2 I56335	clusterin precursor
7	400.5	15.2	439	2 A35744	apolipoprotein J -
8	400.5	15.2	446	2 A42108	clusterin precursor
9	380	14.4	447	2 A27205	clusterin precursor
10	168	6.4	191	2 I48174	sulfated glycoprotein
11	164	6.2	1738	2 T14867	interaptin - slime
12	151	5.7	1526	2 T41522	myosin ii - fission
13	150	5.7	1156	2 B70356	chromosome assembly
14	149	5.7	1388	2 S70633	serine/threonine-s
15	148.5	5.6	1091	2 T34107	hypothetical prote
16	147	5.6	3225	2 I52300	giantin - human
17	147	5.6	3259	1 A56539	giantin - human
18	145.5	5.5	594	1 A44073	C1K1 protein - yea
19	141.5	5.4	1005	2 A44655	hypothetical prote
20	141.5	5.4	3433	1 S28381	utrophin - human
21	140.5	5.3	2166	2 G70163	hypothetical prote
22	140	5.3	1132	2 T00259	hypothetical prote
23	139.5	5.3	2663	1 S28261	centromere protein
24	138	5.2	1084	2 G71329	hypothetical prote
25	135	5.2	1955	2 T30934	myosin-like protei
26	137	5.2	1940	1 S04090	myosin heavy chain
27	136.5	5.2	1410	1 A57013	early endosome ant
28	136	5.2	1388	2 S74245	serine/threonine-s
29	135.5	5.1	1679	2 S48385	hypothetical prote

ALIGNMENTS

RESULT 1

A41386

clusterin precursor [validated] - human

N:Alternate names: apolipoprotein J; complement cytolysis inhibitor SP-40; complement TRPM-2/clusterin protein

N:Contains: clusterin alpha chain; clusterin beta chain

C:Species: Homo sapiens (man)

C>Date: 03-Apr-1992 #sequence_revision 17-Nov-1995 #text_change 08-Dec-2000

C:Accession: S43646; S04662; A41386; A35833; S34036; A53177; A37816; B37816; PL0136;

R:Wong, P.; Taillefer, D.; Lakins, J.; Pineault, J.; Chader, G.; Tenniswood, M.

Eur. J. Biochem. 221, 917-925, 1994

A:Title: Molecular characterization of human TRPM-2/clusterin, a gene associated with

A:Reference number: S43646; MUID:94237156

A:Accession: S43646

A:Molecule type: DNA

A:Residues: 1-449 <WON>

A:Cross-references: GB:M64722; NID:g339972; PIDN:AAB06508.1; PID:g339973

R:Kirsbaum, L.; Sharpe, J.A.; Murphy, B.; d'Apice, A.J.F.; Classon, B.; Hudson, P.;

EMBO J. 8, 711-718, 1989

A:Title: Molecular cloning and characterization of the novel, human complement-associ

A:Reference number: S04662; MUID:89251601

A:Accession: S04662

A:Molecule type: mRNA

A:Residues: 1-449 <KIR>

A:Cross-references: EMBL:X14723; NID:g30250; PIDN:CAA32847.1; PID:g30251

A:Note: parts of this sequence, including the amino end of the mature protein, were c

R:Jenne, D.E.; Tschopp, J.

Proc. Natl. Acad. Sci. U.S.A. 86, 7123-7127, 1989

A:Title: Molecular structure and functional characterization of a human complement cy

tis fluid.

A:Reference number: A41386; MUID:89386692

A:Accession: A41386

A:Molecule type: mRNA

A:Residues: 2-449 <JEN>

A:Cross-references: GB:M25915; NID:g180619; PIDN:AAA35692.1; PID:g180620

R:de Silva, H.V.; Harmony, J.A.K.; Stuart, W.D.; Gill, C.M.; Robbins, J.

Biochemistry 29, 5380-5389, 1990

A:Title: Apolipoprotein J: structure and tissue distribution.

A:Reference number: A35833; MUID:90344779

A:Accession: A35833

A:Molecule type: mRNA

A:Residues: 34-449 <DES>

A:Cross-references: GB:J02908; NID:g178854; PIDN:AAA51765.1; PID:g178855

R:Chiso, J.; Matsubara, E.; Koudinov, A.; Choi-Miura, N.H.; Tomita, M.; Wisniewski, T

Biochem. J. 293, 27-30, 1993

A:Title: Research Communication. The cerebrospinal-fluid soluble form of Alzheimer's

A:Reference number: S34056; MUID:93319521

A:Accession: S34056

A:Molecule type: protein

A:Residues: 228-240, 'X', 242-246; 23-24, 'X', 26-34, 'X', 36-38, 'X', 40-41 <GHI>

R:James, R.W.; Hochstrasser, A.C.; Borghini, I.; Martin, B.; Pometta, D.; Hochstrasse

Arterioscler. Thromb. 11, 645-652, 1991

A:Title: Characterization of a human high density lipoprotein-associated protein, NAL/NA
A:Reference number: A53177; MUID:91230083
A:Accession: A53177
A:Molecule type: protein
A:Residues: 229-242;303-304,'M',306-312,'X',314-317;397-403 <JAM>
R:De Silva, H.V.; Stuart, W.D.; Park, Y.B.; Mao, S.J.T.; Gal, C.M.; Wetterau, J.R.; Busc
J. Biol. Chem. 265, 14292-14297, 1990
A:Title: Purification and characterization of apolipoprotein J.
A:Reference number: A37816; MUID:90354412
A:Accession: A37816
A:Molecule type: protein
A:Residues: 23-46,'H',48-51,'Q' <DE>
A:Note: amino end of the alpha chain
A:Accession: B37816
A:Molecule type: protein
A:Residues: 228-257 <DE2>
A:Note: amino end of the beta chain
R:Choi, N.H.; Mazda, T.; Tomita, M.
Mol. Immunol. 26, 835-840, 1989
A:Title: A serum protein SP40.40 modulates the formation of membrane attack complex of C
A:Reference number: PL0135; MUID:90097955
A:Accession: PL0135
A:Molecule type: protein
A:Residues: 23-37 <CHO>
A:Note: this fragment was isolated from the membrane attack complex SC5b-9
A:Accession: PL0135
A:Molecule type: protein
A:Residues: 228-242 <CH2>
A:Note: this fragment was isolated from the membrane attack complex SC5b-9
R:Hochstrasser, A.C.; James, R.W.; Martin, B.M.; Harrington, M.; Hochstrasser, D.; Pomet
Appl. Theor. Electrophor. 1, 73-76, 1988
A:Title: HDL particle associated proteins in plasma and cerebrospinal fluid: Identification
A:Reference number: S07433; MUID:91265608
A:Accession: S07433
A:Molecule type: protein
A:Residues: 229-240 <HOC>
A:Accession: S07433
A:Molecule type: protein
A:Residues: 24-27,'S',29-33 <HO2>
R:Kriszbaum, L.; Bozas, S.E.; Walker, I.D.
FEBS Lett. 297, 70-76, 1992
A:Title: SP-40.40, a protein involved in the control of the complement pathway, possesses
A:Reference number: A56293; MUID:92201397
A:Accession: A56293
A:Molecule type: protein
A:Residues: 229-240 <HOC>
A:Accession: S07433
A:Molecule type: protein
A:Residues: 229-240 <HOC>
A:Accession: S07433
A:Molecule type: protein
A:Residues: 228-246 <KUN>
A:Experimental source: apolipoprotein A-I-containing lipoproteins, plasma
A:Accession: F54223
A:Molecule type: protein
A:Residues: 23-34,'X',36-37 <KU2>
A:Experimental source: apolipoprotein A-I-containing lipoproteins, plasma
R:Danik, M.; Chabot, J.G.; Mercier, C.; Benabid, A.L.; Chauvin, C.; Quirion, R.; Suh, M.
Proc. Natl. Acad. Sci. U.S.A. 88, 8577-8581, 1991
A:Title: Human gliomas and epileptic foci express high levels of a mRNA related to rat
A:Reference number: I59206; MUID:92020896
A:Accession: I59206
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 61-449 <RES>
A:Cross-references: GB:W74816; NID:g338056; PIDN:AAA60321.1; PID:g338057
R:Duguid, J.R.; Bohmont, C.W.; Liu, N.G.; Tourtellotte, W.W.
Proc. Natl. Acad. Sci. U.S.A. 86, 7260-7264, 1989
A:Title: Changes in brain gene expression shared by scrapie and Alzheimer disease.
A:Reference number: I48174; MUID:89386721
A:Accession: I63132
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 103-168 <RE2>

A:Cross-references: GB:M26639; NID:g338070; PIDN:AAA36609.1; PID:g553644
C:Comment: This protein has been implicated in complement cascade inhibition, membran
11 as in normal brain, in tissues affected by neurodegenerative disease processes, an
C:Comment: This protein may assist in preventing the formation of Alzheimer's disease
C:Genetics:
A:Gene: GDB:CLU; CLI
A:Cross-references: GDB:I25226; OMIM:185430
A:Map position: 8p21-8p21
A:Note: appears to be a single-copy gene; alternative exon usage in 5'-untranslated r
C:Superfamily: clusterin
C:Keywords: apoptosis; complement inhibitor; extracellular protein; glycoprotein; HDL
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-227/Domain: clusterin beta chain #status experimental <BCH>
F:23-227,228-449/Product: clusterin #status experimental <NAT>
F:228-449/Domain: clusterin alpha chain #status experimental <ACH>
F:86,103,145,291,354,374/Binding site: carbohydrate (Asn) (covalent) #status experime
F:102-313,121-295,129-285/Disulfide bonds: #status experimental
F:113-305,116-302/Disulfide bonds: (or 113-302, 116-305) #status experimental
F:317/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 16.5%; Score 434; DB 1; Length 449;
Best Local Similarity 25.8%; Pred No. 7.3e-20;

Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

QY 30 MKPPLLIVFIVLLMLKDSHCAPTWK-----DKTAISENLKSFSEVEIDADEEVKKALT 83

Db 1 MMKTLTLFVGLLL-----TWESGQVLDGDTVSDNELQMSNQSGSVYKNKEIQNAVN 51

QY 84 GIKQKIMMERKEKEHTNLMSTLKCRBQKALKNVQHLBEERLCRESLADSG 143

Db 52 GYKQIKTLIEKTNEBKRTLLSLEAKKKKDALNETRESETKLKLPGVCNETMMALWE 111

QY 144 ECRSCLENNCRITYT--TCQPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIK 202

Db 112 ECKPCLKQTCMKFYARVCRSGSLVGRQLEELNQSPPYFWMGDR-----ISLLEN 165

QY 203 DAQLTQMEDV----FSQLTVDVNSILNRSFNVRMQQEDFDTQFSHFIS--TDTJTEPYF 257

Db 166 DRQQTMLMDVMQDHFSSRASSIIDELFQDRF-----FTREPQDTY--HYLPFSLPHRRPHF 218

QY 258 RPAFSKEPMTKADLE--QCWDIDNFTQLCNFSVSIVSEYITMKLKAIEDLPQDKAP 316

Db 219 F--FPKSRIVRSLMPSPSPYFNFHAMFQPFLEMIHEA-----QOAMDIFHFSAPF 267

QY 317 DHGGLSKMLPQDRLGCLGELQNLRSRCFKFHEKQCKQAOHLSEDC----PDVPALHTBL 372

Db 268 QHPPTFIREGDDRTVCREIRHNSGCLRMKQDKCKREILSVDCSTNNPQAKLRRL 327

QY 373 DEAIRLVNVSNOQYQIILQMTKHLLEDYALVKEKRGQFGVYSELANOAPETEIFNSIQ 432

Db 328 DESLQVAERLTKRYNELLSYQMKMLNTSSLLEQLNEQFNWYSRLANTQGGEDQYLRVT 387

QY 433 VVPRIHGNSIKODETMTDLSLPNFTLKLPLEESAESSNFYGVYVAKALQHF--KEH 491

Db 388 TVAS--HTSDSDVSPGVTEVVVVKLFDSDFITVTPVEVSRKPKFMETVAERKALQETRRKH 446

RESULT 2

150131

clusterin - quail

C:Species: Coturnix coturnix (quail)

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999

C:Accession: 150131

R:Michel, D.; Chatelain, G.; Herault, Y.; Brun, G.

Eur. J. Biochem. 229, 215-223, 1995

A:Title: The expression of the avian clusterin gene can be driven by two alternative

A:Reference number: 150131; MUID:95262670

A:Accession: 150131

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-451 <MIC>

A:Cross-references: EMBL:X80760; NID:g520629; PIDN:CAA56733.1; PID:g520630

C:Genetics:

A:Gene: T64

A:Introns: 29/1; 78/3; 135/3; 275/1; 313/1; 388/3; 450/2

C:Superfamily: clusterin

Query Match 15.9%; Score 418; DB 2; Length 451;

Best Local Similarity 25.7%; Pred. No. 7.3e-19;

Matches 118; Conservative 94; Mismatches 179; Indels 68; Gaps 15;

QY 63 LKSFSEVEIDADEVKKALTGKIMMERKEHTNLMSTLKKREKQEAALKLLNE 122

Db 27 LKLSAAGSKYIDAEVENAINGVKMKTLMKDTSKSKEHQAHLTEETKKKEAAVKLAE 86

QY 123 VQHELEEEERLCRESLADSGEGRSCLENMCRITY--TCOPSSVSKNIERFRKIIYQF 181

Db 87 KEKQAEKQEVNCTNLSLWEECKPCLKHCTCMRVYSKMGSGSLVGROLEEFNRSPPF 146

QY 182 LFPFHENEKDLPISEKLIK-----DAQLTQMEDVFSQTLVDVNSLNFNSFNVMQ 236

Db 147 SIWNGERIDDLDRQORRRFEDLEERFGLMEDGVEDIFQDSTQLYGPAFPFR--- 202

QY 237 EFDQTFQSHFISDTDLTEPY--FFPAFSKEPMTKADLEQCWDIPN-----FFQLFC 285

Db 203 -----TPPGGFREAF--VPPVORVHL-----VPRRLSRELHPFFOHPM 240

QY 286 NFSVSIYESVETITKMLKAIEDLPKQDKAPDH--GGLISKMLP-GODRGLGCGELDONLS 342

Db 241 HGPHRLFOPLFEMTOHMLDG-----CHGAWHPGLGGFATESRNFSTRMVCREIRNSA 294

QY 343 RCKFHEKCKQCNHLSEDC-----PDVPAHLTELDEAIRLVNSVNOQYGOILQWTRKHE 398

Db 295 GCLRMDECEKREILAVDCSQTDPVQSOLREQFEDALRAERFTRRYDILLSAFOAEML 354

QY 399 DTALVEKMRGQGVWSELANQAP-----ETETIFNSIQVVPRIHSGNISKQDETMTD 452

Db 355 NTSLLDQLNRQGVWSELNLTQNGDGLQVTVTSK-----TPNLEDP--SAPADTQVT- 408

QY 453 LSLIPSSNFTLKIPLESSESSNFIYVVAKALQHFKEH 491

Db 409 VOLFDSEPLSLTVPGDISDSDPRFMEIVAEQALQHYKQN 447

RESULT 3

A40018

N:clusterin precursor - dog

C:Species: Canis lupus familiaris (dog)

C:Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 13-Aug-1999

C:Accession: A40018

R:Hartmann, K.; Rauch, J.; Urban, J.; Parczyk, K.; Diel, P.; Pillarsky, C.; Appel, D.; Ha

J. Biol. Chem. 266, 9924-9931, 1991

A:Title: Molecular cloning of gp 80, a glycoprotein complex secreted by kidney cells in

F:227-445/Domain: clusterin alpha chain #status experimental <ACH>

F:86,103,145,277,287,350,370/Binding site: carbohydrate (Asn) (covalent) #status predict

A:Accession: A40018

A:Molecule type: mRNA

A:Residues: 1-445 <HAR>

A:Cross-references: GB:M55251; GB:M38757; NID:g163953; PIDN:AAA30846.1; PID:g163954

A:Note: parts of this sequence, including the amino ends of the alpha and beta chains,

C:Superfamily: clusterin

C:Keywords: apoptosis; disulfide bond; glycoprotein

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-226/Domain: clusterin beta chain #status experimental <BCH>

F:227-445/Domain: clusterin alpha chain #status experimental <ACH>

F:86,103,145,277,287,350,370/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match

Best Local Similarity 15.6%; Score 410.5; DB 2; Length 445;

Matches 117; Conservative 102; Mismatches 192; Indels 79; Gaps 15;

QY 30 MKPPLLVIICLVWLKDSHCAPTWKDKTAISENLKSFSEVEIDAD-----BEVKKALT 83

Db 1 MMKTLALLVGLLL-----TWNGRVLCDAQVSDTELOQMSTEGSKYINKKIKKALK 51

QY 84 GIKOMKIMMERKEHEHTNLMSTLKKREKQEAALKLLNEVQEHLEERLCRESLADSWG 143

Db 52 GVKQIKTLIEQTNEERKSLNLSEAKKKEDALNDTKDSETKLKASQGVNDTMMALWE 111

QY 144 ECRSCLNNCMRIYT--TCOPSSVSKNIERFRKIIYQFLFPFHEDNEKDLPISEKLIK 202

Db 112 ECKPCLQTCMKFVARVCRSGSLVGHQLEEFNLQSSPFYFMNGDR-----IDSLLEN 165

QY 203 DAQLTQMEDVFSQTLVDVNSLNFNSFNVMQOQEFQTFQSHFIS-----DTDLTEPYFF 258

Db 166 DRQOATHADV-----MQDSFNRSASI-----MDQLFQDRFTRPQDTYHSP--F 209

QY 259 PAFSKEPMTKADLEQCWDI-----PNFQLFCNFSVSIYESVETITKMLKAIE--- 307

Db 210 SLFORRPFNPKFIARNIIPFPFQPLNPHDMFQEPFDMIHQA--QOAMDVNLHRIYPHF 268

QY 308 --DLPKQDKAPDHGGLISKMLPGODRGLGCGELDONLSRCFKEHCKCQQAHLSEDC--- 362

Db 269 PIRFPEDE-----NRTVKEIRHNSGCLMKDKCEKCEILSVDCSSN 312

QY 363 --PDVPAHLTELDEAIRLVNSVNOQYGOILQWTRKHELTAYLVKMRGQGVWSELANQA 421

Db 313 NPAQVOLRQELSNLSQIAEKFTKLYDELQSYQEKMENTSSLKQLNEQFSWVSQLANLT 372

QY 422 PETEIIFNSIQVVP--PRIHSGNISKQDETMTDLSILPSSNFTLKIPLESSESSNFIYV 480

Db 373 QSDPPFYLVTVTGSQTSNPNVPGFTKV--VKLFSDPITVMIPAVSRNNPKMETV 430

QY 481 VAKALQHFKE 490

Db 431 AERKALQERYQ 440

RESULT 4

S07714

T64 protein precursor - Japanese quail

C:Species: Coturnix coturnix japonica (Japanese quail)

C:Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 13-Aug-1999

C:Accession: S07714

R:Michel, D.; Gillet, G.; Volovitch, M.; Pessac, B.; Calothy, G.; Brun, G.

Oncogene Res. 4, 127-136, 1989

A:Title: Expression of a novel gene encoding a 51.5 kD precursor protein is induced b

A:Reference number: S07714; MUID:89239492

A:Accession: S07714

A:Molecule type: mRNA

A:Residues: 1-449 <MIC>

A:Cross-references: EMBL:X15825; NID:g62594; PIDN:CAA33823.1; PID:g62595

C:Superfamily: clusterin

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-449/Product: T64 protein #status predicted <MAT>

Query Match

Best Local Similarity 15.5%; Score 408; DB 2; Length 449;

Matches 117; Conservative 92; Mismatches 179; Indels 68; Gaps 15;

QY 63 LKSFSEVEIDADEVKKALTGKIMMERKEHTNLMSTLKKREKQEAALKLLNE 122

Db 27 LKLSAAGSKYIDAEVENAINGVKMKTLMKDTSKSKEHQAHLTEETKKKEAAVKLAE 86

QY 123 VQHELEEEERLCRESLADSGEGRSCLENMCRITY--TCOPSSVSKNIERFRKIIYQF 181

Db 87 KEKQAEKQEVNCTNLSLWEECKPCLKHCTCMRVYSKMGSGSLVGROLEEFNRSPPF 146

QY 182 LFPFHEDNEKDLPISEKLIK-----DAQLTQMEDVFSQTLVDVNSLNFNSFNVMQ 236

Db 147 SIWNGERIDDLDRQORRRFEDLEERFGLMEDGVEDIFQDSTQLYGPAFPFR--- 202

QY 237 EFDQTFQSHFISDTDLTEPY--FFPAFSKEPMTKADLEQCWDIPN-----FFQLFC 285

```
Db 203 -----TPPFGGFEAF-VPPVQVRVHL-----VPRRRLSRELHPFQHPM 240
Qy 286 NFSVSIYESVETITKMLKAIEDLPKQADPH--GGLISKMLP--GQDRGLCGELDONLS 342
Db 241 HGFHRLFOPLFEMTQHMLDG-----GHGAWHEPLGGFATESRNFSTDRMVCREIRNSA 294
Qy 343 RCFKFEKCKQKQAHLSDEC-----PDVPALHTLEDEARLVNVSQQYQGIQLQMTKKHLE 398
Db 295 GGLMRDECEKREILAVDCSDPVPQSOLRQFEDALRAERFTRRYDDLLSAFQAEML 354
Qy 399 DTAIVLEKMRGQFGWVSELANQAP-----ETEILFNSIQVVPRIHEGNISKQDETMWD 452
Db 355 NTSLLDQLNRFGWVSRIGNLTQNGDGLQVTTVFESK-----TPNLEDP--SAPADTQVT- 408
Qy 453 LSLIFSNFTLKIPLEESAESSNFYGYVAKALQHF 488
Db 409 VQLFDSLEPLSLVPGDISWDDPRFMEIVPEQALQHY 444

RESULT 5
A40714
Clusterin precursor - mouse
N:Alternate names: sulfated glycoprotein 2
N:Contains: clusterin alpha chain; clusterin beta chain
C:Species: Mus musculus (house mouse)
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 13-Aug-1999
C:Accession: A40714; JN0699
R:French, L.E.; Chonn, A.; Ducrest, D.; Baumann, B.; Belin, D.; Wohlwend, A.; Kiss, J.Z.
J. Cell Biol. 122, 1119-1130, 1993
A:Title: Murine clusterin: molecular cloning and mRNA localization of a gene associated
A:Reference number: A40714; MUID:93359508
A:Accession: A40714
A:Molecule type: mRNA
A:Residues: 1-448 <FRE>
A:Cross-references: GB:L08235; NID:g192596; PIDN:AAA37422.1; PID:g192597
R:Lee, K.H.; Ji, Y.M.; Lim, H.M.; Lee, S.C.; You, K.H.
Biochem. Biophys. Res. Commun. 194, 1175-1180, 1993
A:Title: Molecular cloning and sequencing of sulfated glycoprotein-2 cDNA from testis of
A:Reference number: JN0699; MUID:93356785
A:Accession: JN0699
A:Molecule type: mRNA
A:Residues: 1-448 <LEE>
A:Note: the authors translated the codon CGC for residue 66 as Ala
C:Comment: This protein is involved in many different cellular processes other than those
C:Superfamily: clusterin
C:Keywords: apoptosis; disulfide bond; glycoprotein; spermatogenesis
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-226/Domain: clusterin beta chain #status predicted <BCH>
F:227-226,227-448/Product: clusterin #status predicted <MAT>
F:227-448/Domain: clusterin alpha chain #status predicted <ACH>
F:102,144,290,327,353,373/Binding site: carbohydrate (Asn) #status predicted

Query Match 15.3%; Score 403; DB 2; Length 448;
Best Local Similarity 23.4%; Pred. No. 6.3e-18;
Matches 112; Conservative 109; Mismatches 196; Indels 62; Gaps 14;

Qy 34 LVFIVCLLWLKDSHCAPTWKOKTAISENLKSFSEVEIDA-----DEEVKALTGIQK 87
Db 3 ILLLCVALLI-----WDNGVILGEQEVSDNELQELSTQGSRYINKEIQNAVQGVKH 54
Qy 88 MKIMMERKEHTNLMSTLKKREEKQKALKLLNEVQHELEERLCRESLADSGECRS 147
Db 55 IKTLIEKTNAEKRLNSLEEAKKKEDALEDTROSEMKLKAFPEVCNETMMALWEECKP 114
Qy 148 CLENNCMRIYV--TCQPSWSVKNKIRFRFKIYQFLFPFHEDNEKDLPISEKLEIKDAQL 206
Db 115 CLKHTCKFYARVCRSGSLVGQOLEEFLNQSSPFYFMNGDR-----IDSLESDRQ 168
Qy 207 TQ-----MEDVFSQLTVDVNSLNRSNFVRMQQEQFDQTFQSHFISDTLTPEYFPF 259
Db 169 SQVLDAMQDSFARASGIIDTLFQDRF--FAR-----ELHDPHYFS-----PIGPPH 214
```

```
Qy 260 ---AFSEKPMTKADLEQC-WDIPNFFQLFCNFSVSIYESVETITKMLKAIEDLPKQDKA 315
Db 215 PHFLYPKRLSLRSLMSPSHYGPSPFNMFQPFEMIHQAQAMDVOLHSPAFQFPDVD-- 272
Qy 316 PDHGLISKMLPGQ--DRGLCGELDONLSRCFKFEKCKQKQAHLSDEC-----PDVPALHT 370
Db 273 -----FLRGEEDRTVCKEIRNRSTGCLMKKQCEKQCEILSVDCSTNNPAQANLRQ 324
Qy 371 ELDEAIRLVNVSQQYQGIQLQMTKKHLEDYALVLEKMRGQFGWVSELANQAPETEIFNS 430
Db 325 ELNDSIQVAERLIEQYKELLQSFQSKMLNTSLLEQLNDQFNWVSQLANLTQGEDKYLYR 384
Qy 431 IQVPRIHENISKQDETMMDLSILPSSNFTLKLPLEESAESSNFYGYVAKALQHF 489
Db 385 VSTV--TTHSSDSEVPVSRVTEVVVYVLFDSDPITVLPPEVSKDNPKFMDTVAEKALQEYR 442

RESULT 6
156335
apolipoprotein J - mouse
C:Species: Mus sp. (mouse)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 13-Aug-1999
C:Accession: I56335
R:Jordan-Starck, T.C.; Lund, S.D.; Witte, D.P.; Aronow, B.J.; Ley, C.A.; Stuart, W.D.
J. Lipid Res. 35, 194-210, 1994
A:Title: Mouse apolipoprotein J: characterization of a gene implicated in atherosclerosis
A:Reference number: I56335; MUID:94223204
A:Accession: I56335
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-448 <RES>
A:Cross-references: GB:S70244; NID:g546504; PIDN:AAB30623.1; PID:g546505
C:Genetics:
A:Gene: ApoJ
C:Superfamily: clusterin

Query Match 15.3%; Score 402; DB 2; Length 448;
Best Local Similarity 24.1%; Pred. No. 7.3e-18;
Matches 115; Conservative 108; Mismatches 201; Indels 54; Gaps 15;

Qy 30 MKPPLLVFIVCLLWLKDSHCAPTWKDKTAISEN-LKSFSEVEIDADEVKALTGIKQM 88
Db 1 MKILLLCVAMLLIW--DNGWV---LGEQEVSDNEIQELSTQGSRYINKEIQNAVQGVKH 55
Qy 89 KIMMERKEHTNLMSTLKKREEKQKALKLLNEVQHELEERLCRESLADSGECRS 148
Db 56 KTLIEKTNAEKRLNSLEEAKKKEDALEDTROSEMKLKAFPEVCNETMMALWEECKP 115
Qy 149 LENNCMRIYV--TCQPSWSVKNKIRFRFKIYQFLFPFHEDNEKDLPISEKLEIKDAQLT 207
Db 116 LKHTCKMFYARVCRSGSLVGQOLEEFLNQSSPFYFMNGDR-----IDSLESDRQ 169
Qy 208 Q-----MEDVFSQLTVDVNSLNRSNFVRMQQEQFDQTFQSHFISDTLTPEYFPF 259
Db 170 QVLDAMQDSFARASGIIDTLFQDRF--FAR-----ELHDPHYFS-----PIGPPH 215
Qy 260 ---AFSEKPMTKADLEQC-WDIPNFFQLFCNFSVSIYESVETITKMLKAIEDLPKQDKAP 316
Db 216 HFLYPKSLRSLMSPSHYGPSPFNMFQPFEMIHQAQAMDVOLHSPAFQFPDVD-- 272
Qy 317 DHGGLISKMLPGQ--DRGLCGELDONLSRCFKFEKCKQKQAHLSDEC-----PDVPALHT 371
Db 273 -----FLUREGDDRTVCKEIRNRSTGCLMKKQCEKQCEILSVDCSTNNPAQANLRQ 325
Qy 372 LDEAIRLVNVSQQYQGIQLQMTKKHLEDYALVLEKMRGQFGWVSELANQAPETEIFNS 431
Db 326 LNDLSQVAERLIEQYKELLQSFQSKMLNTSLLEQLNDQFNWVSQLANLTQGEDKYLYR 385
Qy 432 QVVPRIHENISKQDETMMDLSILPSSNFTLKLPLEESAESSNFYGYVAKALQHF 489
Db 386 STV--TTHSSDSEVPVSRVTEVVVYVVKLFDSDPITVLPPEVSKDNPKFMDTVAEKALQEYR 442
```

RESULT 7

A35744
clusterin precursor - bovine
N:Alternate names: complement cytolysis inhibitor; glycoprotein III
N:Contains: clusterin alpha chain; clusterin beta chain
C:Species: Bos primigenius taurus (cattle)
C:Date: 05-Oct-1990 #sequence_revision 05-Oct-1990 #text_change 13-Aug-1999
C:Accession: A35744
R:Palmer, D.J.; Christie, D.L.
J. Biol. Chem. 265, 6617-6623, 1990
A:Title: The primary structure of glycoprotein III from bovine adrenal medullary chromaffin granules
A:Reference number: A35744; MUID:90216681
A:Accession: A35744
A:Molecule type: mRNA
A:Residues: 1-439 <PAL>
A:Cross-references: GB:J05391; NID:g163114; PID:AAA30554.1; PID:g163115
A:Note: parts of this sequence, including the amino ends of the alpha and beta chains, were determined from complementary DNA
C:Superfamily: clusterin
C:Keywords: apoptosis; disulfide bond; extracellular protein; glycoprotein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-221/Domain: clusterin beta chain #status experimental <BCH>
F:221-222/Domain: clusterin #status experimental <MAT>
F:222-439/Domain: clusterin alpha chain #status experimental <ACH>
F:80,97,139,283,320,346,366/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 15.2%; Score 400.5; DB 2; Length 439;
Best Local Similarity 24.5%; Pred. No. 8.8e-18;
Matches 114; Conservative 97; Mismatches 183; Indels 71; Gaps 15;

QY 52 TWKDKTAISE-NLKSFEVEGEIDAEVKKALGTGKIMMERKEKHTLMSTLKKCR 110
Db 13 SWEGSAISKELOEMSTEGSKYVKNKEIKALKEVQIKTQIEQNEERKLLLSLEAK 72
QY 111 EKEQALKLLNEVOEHLLEERLCRESLADSGWGCRCLENNCMRIYT-TCQPSWSSVK 169
Db 73 KKKEDALNDRSDENKLGASGVGCVNETWTALWEECKPCLKOTCMKFYARVCRSGSLGVGH 132
QY 170 KIERFRKRYQFLPPFHEDNEKOLPISEKLEIKDAQLTOMEDVFSQLTVDVNSLFRSEN 229
Db 133 QLEEFLLNQSPFYFWINGDR-----IDSLMENDREQSHVMDV-----MEDSFTRASS 179
QY 230 VFRQMQQEFDTQSHFI-----SDTDLTEPY-----FF-----PAPSKPMTKADLEQC 274
Db 180 I-----MDLFQDRFFLRPPQDTQYSPFSGRSLFFNPKSRFARNVMPFLLEPF 232
QY 275 WDPNPFQFCNFSVSIYESVSETITKMLKAIEDLPKQDAPDHGGGLISKMLPCQDRGLC 334
Db 233 -----NPHDVFPFP-----YDMLHQAQAQMDAHLQ-----RTPHFPTM-EFTENNORTVC 277
QY 335 GELDONLSKCFKHEKQKCOAHLSEDC-----PDVPAHLTELDEAIRLVNVSNOQYQGIL 390
Db 278 KEIRHNSGTGLRMKQDCEKQCEILEVDSCASNPTQTLRLQOLNASLQAEKFSRLYDQL 337
QY 391 QMTKRKHLDEPAYLVEKMRGQGVWSELANAQAPETE-----IIFNSIQVVPRIHEGNISK 444
Db 338 QSQOQKMLNATSALEKQNEQFTWVSQLANLTQSDQHQYLVQFTVNSHNSDPSPSG----- 393
QY 445 QDETMTDLSILPSSNFTLKIPLEESAESSNFYGVVAKALQHF 489
Db 394 -----LTKVIVKLFNSFPITVTPQEVSSPNFMENVAEKALQYR 433

RESULT 8

A42108
clusterin precursor - pig
N:Alternate names: complement cytolysis inhibitor; CP40 protein
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
C:Accession: A42108; JC5535; PC4475
R:Dlemer, V.; Hoyle, M.; Baglioni, C.; Millis, A.J.
J. Biol. Chem. 267, 5257-5264, 1992

A:Title: Expression of porcine complement cytolysis inhibitor mRNA in cultured aortic
A:Reference number: A42108; MUID:92184774
A:Accession: A42108
A:Molecule type: mRNA
A:Residues: 1-446 <DIE>
A:Cross-references: GB:M84639; NID:g164408; PID:AAA31013.1; PID:g164409
A:Experimental source: aortic smooth muscle cells
A:Note: sequence extracted from NCBI backbone (NCBIN:87354, NCBI:87356)
R:Ogawa, S.; Ishibashi, Y.; Sakamoto, Y.; Kitamura, K.; Kubo, M.; Sakai, T.; Inoue, K.
Biochem. Biophys. Res. Commun. 234, 712-718, 1997
A:Title: The glycoproteins that occur in the colloids of senescent porcine pituitary
A:Reference number: JC5535; MUID:97318844
A:Accession: JC5535
A:Molecule type: DNA
A:Residues: 1-446 <OGA>
A:Accession: PC4475
A:Molecule type: protein
A:Residues: 58-66;68-77;229-247;249-251;408-436 <OG2>
A:Experimental source: pituitary gland
C:Superfamily: clusterin
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-227/Domain: clusterin beta chain #status predicted <BCH>
F:23-227,228-446/Product: clusterin #status predicted <MAT>
F:228-446/Domain: clusterin alpha chain #status predicted <ACH>

Query Match 15.2%; Score 400.5; DB 2; Length 446;
Best Local Similarity 24.4%; Pred. No. 9e-18;
Matches 118; Conservative 98; Mismatches 196; Indels 71; Gaps 16;

QY 34 LLVFIIVCLWLKDSHCAPTWK-----DKTAISENLKSFSEVEGEIDAEVKKALGTGK 86
Db 4 LLLLVGLLL-----TWENGFWLGLDKAISDELQEMSTEGSKYVKNKEIKALKEV 54
QY 87 QMIMMERKEKHTLMSTLKKCREKQEAALKLLNEVOEHLLEERLCRESLADSGWGC 146
Db 55 QIKTLIEQSNEEKRSLSSLEAKKKEDALNDRDTETKLGKSGQGLCNETMMALWEECK 114
QY 147 SCLENNCMRIYT-TCQPSWSSVKNIETFRKTYQFLPPFHEDNEKDLPISEKLEIKDAQ 205
Db 115 PCLQTCMKFYARVCRSGGLVGHQLEEFLLNQSPFYFWINGDR-----IDSLMENDRQ 168
QY 206 LTOMEDVFSQLTVDVNSLFRSENFRQMQQEFDTQSHFIS-----DSDLTEPY----- 256
Db 169 QSHVMDI-----MEDSFNRASNI-----MDLFQDRFFNRFPFDQTFQFPFGSSHR 214
QY 257 ---FFPAFSKPMTKADLEQCWDIPNFFQFCNFSVSIYESVSETITKMLKAIEDLPKQD 313
Db 215 GSLFFNPKSRFARNIMPFPLFTDL-NYHDMFQFP-----FDMIHQAQAQMDAHLHRIYH- 268
QY 314 KAPDHGGLISKMLP--GODRGLGGLDQNLSCFKHEKQKCOAHLSEDC-----PDVPA 367
Db 269 -FPEAG-----VPENSNDRVAVCKEIRHNSGTGLRMKQDCEKREILSDCSASNSSQ 321
QY 368 LHTELDEAIRLVNVSNOQYQGLQMLTKHLEDTAYLVEKMRGQGVWSELANAQAPETII 427
Db 322 LRQELYTSIQMAEKFSKLYDQLLQSYQKMLNTSSLKQLNEQFSWVSQLANLTQNDRY 381
QY 428 FNSIQVVPRIHEGNISKQDETMMTDLISLPSSNFTLKIPLEESAESSNFYGVVAKALQH 487
Db 382 YLQVTTV-NSHGSDPSVPSGLTKVVVVKLFDSYPITLIIPQEV--DPKFMETVAEALQ 438
QY 488 FKE 490
Db 439 YRQ 441

RESULT 9

A27205
clusterin precursor - rat
N:Alternate names: SGP-2; sulfated glycoprotein 2; TRPM-2
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 13-Aug-1999

C:Accession: A45890; S18491; A45415; B31575; A31575; A27205
R:Collard, M.W.; Griswold, M.D.
Biochemistry 26, 3297-3303, 1987
A:Title: Biosynthesis and molecular cloning of sulfated glycoprotein 2 secreted by rat S
A:Reference number: A45890; MUID:88000523
A:Accession: A45890
A:Molecule type: mRNA
A:Residues: 1-447 <CO2>
R:Betts, J. S.; Hlipakka, R.A.; Gilna, P.; Liao, S.
Biochem. J. 257, 293-296, 1989
A:Title: Identification of an androgen-repressed mRNA in rat ventral prostate as coding
A:Reference number: S18491; MUID:89149740
A:Accession: S18491
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-186, 'D', 188-447 <BET>
A:Cross-references: EMBL:X13231; NID:g57240; PIDN:CAA31618.1; PID:g57241
R:Wong, P.; Pineault, J.; Lakins, J.; Taillefer, D.; Leger, J.; Wang, C.; Tenniswood, M.
J. Biol. Chem. 268, 5021-5031, 1993
A:Title: Genomic organization and expression of the rat TRPM-2 (clusterin) gene, a gene
A:Reference number: A45415; MUID:93186813
A:Accession: A45415
A>Status: preliminary
A:Molecule type: DNA; mRNA
A:Residues: 1-447 <WON>
A:Experimental source: blood
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:126803, NCBI:P:126805)
R:Cheng, C.Y.; Chen, C.L.C.; Feng, Z.M.; Marshall, A.; Bardin, C.W.
Biochem. Biophys. Res. Commun. 155, 398-404, 1988
A:Title: Rat clusterin isolated from primary Sertoli cell-enriched culture medium is sul
A:Reference number: A90146; MUID:88326333
A:Accession: B31575
A:Molecule type: protein
A:Residues: 22-51 <CHE>
A:Note: amino end of the beta chain
A:Accession: A31575
A:Molecule type: protein
A:Residues: 227-256 <CH2>
A:Note: amino end of the alpha chain
C:Superfamily: clusterin
C:Keywords: apoptosis; disulfide bond; glycoprotein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-226/Domain: clusterin beta chain #status experimental <BCH>
F:227-447/Domain: clusterin alpha chain #status experimental <MAT>
F:227-447/Domain: clusterin alpha chain #status experimental <ACH>
F:102,144,290,327,353,373/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.4%; Score 380; DB 2; Length 447;
Best Local Similarity 23.5%; Pred. No. 1.7e-16;
Matches 113; Conservative 103; Mismatches 197;

QY 34 LVFIVICLLWLKDSHCAPTKWKTATSENLSKFSVEGEID-----ADEEVKALTGIKQ 87
DB 4 LLLCVALL-----TDWNGMVLGEQEFSDNEIQLSTQGSRYRYEIONAVQGVKH 54
QY 88 MKTMRKEKHTNLSMTLKKCKREEKQKALKILNEVQEHLEEEERLCRESLADSWGECRS 147
DB 55 IKTLIEKTNARKSLNLSLEAKKKEGALDDTDFDSEMLKAPFVCNETMMALWEECKP 114
QY 148 CLENNCKRIYV-TCQPSWSVKNKIERFFKRIYQVLPFPHEKNEKDLPTSEKLIKEDQAL 206
DB 115 CLKHTCKMFARVCRSGSLVGRQLEEFNLQSSPFYFWMNGDR-----IDSLLSDRQ 168
QY 207 TQ-----MEDVFSQITVDVNSLFRNSNVFRMQOEEDQTFQSHFISDTDLTEPYFFP--- 259
DB 169 SQVLDAQMSDFTRASGLIHFLFDQRF--FTHEPQDI-----HHF-----SPMGFPK 214
QY 260 ---AFSKEPTKADLEQCWDIP--NFFOLFCNFSVIVSVSEITKMLKAIEDLPKQDKA 315

DB 215 PHFLYPKSRLVRSMLPSLHSHYGLPSLHNFQFFDMIHQAQQAMDVQLHSPALQFPDVFDFL 274
QY 316 PDHGLISKMLPGQ--DRGLGCGELDONLSRCFKFHEKCKQCOAHLSEDC-----PDVPALHT 370
DB 275 KE-----GEDDPTVCKEIRHNSGCLMKMGQCEKQCEILSVDCSTNNPAQANLRQ 324
QY 371 ELDEAIRLVNVSNOQYGOILQMTKRKHLEDYALVKEKMRGQFGWVSELAN--QAPETELIF 428
DB 325 ELNDSLOVAERLTOQYNELLHSLQSKMLNTSSLLEQLNDQFTWVSQLANLTQGDQDYLRV 384
QY 429 NSIQVVPRIHEGNISKQDETMTDLSIIPSSNFTLKIPLEESAESNFIQVYVAKALQHF 488
DB 385 STVTT-----HSDSEVPVSRVTEVVVKLFDSPTITWLPVEVSKDNPKFMDIVAELQEQY 440
QY 489 K 489
DB 441 R 441
RESULT 10
148174
sulfated glycoprotein 2 - golden hamster (fragment)
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 13-Aug-1999
C:Accession: I48174
R:Duguid, J.R.; Bohmont, C.W.; Liu, N.G.; Tourtellotte, W.W.
Proc. Natl. Acad. Sci. U.S.A. 86, 7260-7264, 1989
A:Title: Changes in brain gene expression shared by scrapie and Alzheimer disease.
A:Reference number: I48174; MUID:89386721
A:Accession: I48174
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-191 <RES>
A:Cross-references: GB:M26640; NID:g191450; PIDN:AAA37102.1; PID:g191451
C:Superfamily: clusterin

Query Match 6.4%; Score 168; DB 2; Length 191;
Best Local Similarity 35.1%; Pred. No. 0.0012;
Matches 33; Conservative 22; Mismatches 35; Indels 4; Gaps 1;

QY 330 DRGLGCGELDONLSRCFKFHEKCKQCOAHLSEDC-----PDVPALHTDELDEAIRLVNVSNOQ 385
DB 69 DRVCKEIRHNSGCLMKMGQCEKQCEILSVDCSANNPAQAHRLQELNDSLQVAERLTOR 128
QY 386 YCQILQMTKRKHLEDYALVKEKMRGQFGWVSELAN 419
DB 129 YNELLHSLQTKMLNTSSLLEQLNEQFNWVSQLAN 162

RESULT 11
T14867
Interaptin - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C:Accession: T14867
R:Rivero, F.J.; Kuspa, A.; Brokamp, R.; Matzner, M.; Noegel, A.A.
J. Cell Biol. 142, 735-750, 1998
A:Title: Interaptin, an actin-binding protein of the alpha-actinin superfamily in Dic
ts.
A:Reference number: Z18248; MUID:98365468
A:Accession: T14867
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1738 <RIV>
A:Cross-references: EMBL:AF057019; NID:g3549260; PID:g3549261; PIDN:AAC34582.1
C:Genetics:
A:Gene: abpd
A:Introns: 173/2; 1680/1

Query Match 6.2%; Score 164; DB 2; Length 1738;
Best Local Similarity 21.1%; Pred. No. 0.029;

[illegible]

```
Db 422 KEDINKLISREKIKKEKEQEIKRKAIAKKKEEELRNLTQELNIYEKRLSEVRKKL 481
QY 398 E-----DTAYLVKMRQFGWVSELAN-QAPE-----423
Db 482 EYLVKKEGATERVRSFSDVSDFKDKIGVYGSVELIRVKNDEHITAIEVAGGRLKFI 541
QY 424 ----TEIIFNSIQVVPRIHEGNIS 443
Db 542 VVEDEEVAKECIQLAKRMNLGRFS 565

RESULT 14
S70633
serine/threonine-specific protein kinase (EC 2.7.1.1.); Rho-associated - bovine
N:Alternate names: Rho-associated protein kinase
C:Species: Bos primigenius taurus (cattle)
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 24-Sep-1999
R:Accession: S70633; S77694
R:Matsumi, T.; Amano, M.; Yamamoto, T.; Chihara, K.; Nakafuku, M.; Ito, M.; Nakano, T.; O
EMBO J. 15, 2208-2216, 1996
A:Title: Rho-associated kinase, a novel serine/threonine kinase, as a putative target fo
A:Reference number: S70633; MUID:96208507
A:Accession: S70633
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1388 <MAT1>
A:Cross-references: EMBL:U36909; NID:g1326077; PIDN:AA048567.1; PID:g1326078
A:Accession: S77694
A:Molecule type: protein
A:Residues: 1-18;30-34;36-44;58-64;133-140;248-252;291-295;327-347;350-360;366-3
-1070 <MAT2>
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase C zinc
C:Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase
F:90-354/Domain: protein kinase homology <KIN>
F:98-106/Region: protein kinase ATP-binding motif
F:1261-1315/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 5.7%; Score 149; DB 2; Length 1388;
Best Local Similarity 16.7%; Pred. No. 0.2;
Matches 95; Conservative 96; Mismatches 166; Indels 212; Gaps 20;

QY 54 KDKTAISENLKSFSEGEIDAEYVKALTGIKOMKI-----MMERKEKEHTNLM 104
Db 780 KQKVLNEDVRNLT-----LKIEQETQKRLTQNDLKMQTQVNTLKMSEKQLKQENHLL 835
QY 105 TLK-----KCRKEQKALLLNEVQHLDEE-----RLCR 135
Db 836 EMKMSLEKQNAELRKERQADQGMKELQDLAEQYFSTLYKTQVRELKECECEKTKLCK 895
QY 136 E-----SLADSWGCRSCLNENCRITYTCOPSSSVKNKIERFRKIYQFLPFPHEDN 189
Db 896 ELQKKQKQELQDE-----RSLAAQLTILTKRQSEQLARSTAEQYSDLEKE 942
QY 190 E--KDLPISE-----KLIKDAQLQTMEDVFSOLTVDVNSLNFNSFNRVFMQOEFDQ 240
Db 943 KIMKELEIKEMMARHQELTEKDNTASLEETNRLTSDVANLANKEEELNNKKEAQEQ 1002
QY 241 TQSHFISDTDLTEYFPFAPSKPEMTADLQOCWDIPNFOLFQCNFSVSIYESVETIT 300
Db 1003 LSR-----LKDEEISAAAIKAQOEKQLLTERTLK-----TQAVN 1036
QY 301 KMLKAIE-----DLPKODKAPDHGGLISKMLPGQDRGLCGEL-----DQNLRCF 345
Db 1037 KLAETMNRKEPVKRGNDTIVRKEK-----ENRKLHMLKLSREKLTQOMI 1082
QY 346 KPFKEKQCKQAHLSDCPDVPALHTDELAIPLVNSVNSQYQGLQMTKRKHLE-----398
Db 1083 KYQKELNEMQAQIAEE-----SQIRLEQMTLDSKDSIEQLRSOLQALHIGLSSSIG 1138
QY 399 -----DTAYLVKMRG-----QFGVSEL-----417
Db 1139 PGDTADDGFPESRLEGLWLSLPVRNNTKFGWKVYIVTSKKILFLFYDSEQDKQSNPYM 1198
```

```
QY 418 -----ANOAPETELIFNSIQVVPRI-----HEGNISKODETMTDLSILPSSNFT 462
Db 1199 VLDIDKLFHVRPVTQDVRADAKELPRIFQILYANEGESKKEQE-----1243
QY 463 LKPILEESAESSNFI---GYVVAKALQHF 488
Db 1244 --FPVEPVGEKSNYICKHGHEFIPTLYHF 1270

RESULT 15
T34107
hypothetical protein C18C4.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
R:Accession: T34107
R:Gattung, S.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid C18C4.
A:Reference number: Z21478
A:Accession: T34107
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1091 <GAT>
A:Cross-references: EMBL:U55369; PIDN:AA025825.1; GSPDB:GN00023; CESP:C18C4.5
A:Experimental source: strain Bristol N2; clone C18C4
C:Genetics:
A:Gene: CESP:C18C4.5
A:Map position: 5
A:Introns: 38/3; 95/3; 179/2; 259/3; 301/3; 419/2; 573/3; 613/1; 875/3; 920/2; 959/2;

Query Match 5.6%; Score 148.5; DB 2; Length 1091;
Best Local Similarity 20.3%; Pred. No. 0.16;
Matches 104; Conservative 100; Mismatches 184; Indels 125; Gaps 25;

QY 54 KDKTAISENLKSFSEGEID-----ADEVKKALTGIKOMK--IMMERKEKEHTNLM 103
Db 283 KDNAIL--NVOLREKDGKIDRIQVDLLAAESRAQAQAEEDVRDMKERIITSKDDSDNNLL 341
QY 104 -STLKKCKREEQKALKLNEVQHEELERCLRESLADSWGECRSCLNENCMR-----IY 157
Db 342 QDELRLRTEEKYQQAQKKIENDETIKQETQIRD--LGRSLDEAKKQLQKMSQRNEEVA 400
QY 158 TTCQPSWSVSNKIERFRKIYQFLPFPHEDNEKDLPISEKLEIKDAQLTQ-----208
Db 401 RQGEDSARSMBEEKATKEEIKKLKSOVLQQLQEDLEQLKKRVQ---ELTEQRKVLESKA 457
QY 209 -MEDVFSOLTVDVNSLNFNSFNRVFMQOEQF-----DQTFQ-----243
Db 458 SVADEFGTLMSSLSL--REEN--RQYEEETRSLQTNIRTLQDEVYQHQAITEMKNRAE 513
QY 244 -----SHFIS-----DTOLTEPYFFPAFKEPTMKADLEOCWMDIPNFFOLFQCNFS 288
Db 514 KAEVIEKENHRVQNASSSHDADITRLENEXTQMEALEKADQEKQAIRE-----ASES 568
QY 289 VSIYE-----SVSETITKMLKAIEDLPKODKAPDHGGLISKMLPGQDRGLCGELQNLNRC 344
Db 569 VRVMKREMTASITSDRIOQLSKE--KVDS-----LTRELESSRRRM--EQLEDQTKF 618
QY 345 KPFKEKQCKQAHLSDCPDVPALHTDELDAIRLVNSVNSQYQGLQMTKRKHLEDTAYLV 404
Db 619 LGSHD---ETKAEWMKD-----LHEAQDETEKLTNQAQLKSKNETLTTE--LEDSONLC 668
QY 405 EKMKGQFGWV----SELANQAPETETIIFNSIQVVPRIHEGNISKODETMTDLSILPSSN 460
Db 669 ERLKAQYKADKDYEEKTKVQLREAEDLADRLOAA-QILSGNVESKFSQDKQE-----719
QY 461 FTLPKIPLEESAESSNFIQYVVAKALQHFKEHFK 493
Db 720 --SKIEMERILDNHN-----KELEKIREELK 743
```

Search completed: July 2, 2002, 11:52:00
Job time: 223 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: ^ July 2, 2002, 11:43:54 ; Search time 24 Seconds
(without alignments)
485.459 Million cell updates/sec

Title: US-09-722-544A-4
Perfect score: 2529
Sequence: 1 MRTWDSNGNMKPLLVFI.....FIGYVAKALQHFKEHFKTW 477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2487	98.3	495	US-09-268-992-2	Sequence 2, Appli
2	2464	97.4	466	US-09-268-992-64	Sequence 64, Appli
3	2377	94.0	477	US-09-268-992-4	Sequence 4, Appli
4	2350	92.9	446	US-09-268-992-51	Sequence 51, Appl
5	1755.5	69.4	465	US-09-268-992-49	Sequence 49, Appl
6	1639	64.8	466	US-09-268-992-39	Sequence 39, Appl
7	1558	61.6	450	US-09-268-992-41	Sequence 41, Appl
8	1277	50.5	374	US-09-268-992-43	Sequence 43, Appl
9	1272.5	50.3	373	US-09-268-992-45	Sequence 45, Appl
10	776	30.7	208	US-09-268-992-67	Sequence 67, Appl
11	145	5.7	1388	US-08-685-576-1	Sequence 1, Appli
12	141	5.6	1388	US-08-685-576-4	Sequence 4, Appli
13	130.5	5.2	2482	US-08-328-254-6	Sequence 6, Appli
14	128.5	5.1	3248	US-08-353-700-1	Sequence 1, Appli
15	128.5	5.1	3248	PCT-US95-16216-1	Sequence 1, Appli
16	125.5	5.0	1354	US-08-685-871-2	Sequence 2, Appli
17	121	4.8	828	US-08-993-228-21	Sequence 21, Appl
18	119.5	4.7	3111	US-08-460-309-4	Sequence 4, Appli
19	119.5	4.7	3111	US-08-125-077-4	Sequence 4, Appli
20	118.5	4.7	1618	US-07-853-913-4	Sequence 4, Appli
21	118	4.7	1886	US-08-938-105-3	Sequence 3, Appli
22	118	4.7	2154	US-08-841-349-4	Sequence 4, Appli
23	117	4.6	1098	US-08-923-992A-8	Sequence 8, Appli
24	117	4.6	1104	US-08-923-992A-4	Sequence 4, Appli
25	117	4.6	1939	US-09-310-187A-1	Sequence 1, Appli
26	117	4.6	2052	US-09-045-201A-2	Sequence 2, Appli
27	116.5	4.6	816	US-08-533-306A-6	Sequence 6, Appli

28	116.5	4.6	816	2	US-08-742-923A-6	Sequence 6, Appli
29	116.5	4.6	1057	4	US-09-541-782-10	Sequence 10, Appl
30	116	4.6	1164	4	US-08-923-992A-2	Sequence 2, Appli
31	115.5	4.6	885	2	US-08-533-306A-4	Sequence 4, Appli
32	115.5	4.6	885	2	US-08-742-923A-4	Sequence 4, Appli
33	115.5	4.6	1786	4	US-08-973-462-8	Sequence 8, Appli
34	114.5	4.5	452	2	US-08-686-599A-18	Sequence 18, Appl
35	114.5	4.5	493	2	US-08-686-599A-5	Sequence 5, Appli
36	114.5	4.5	493	2	US-08-686-599A-16	Sequence 16, Appl
37	114	4.5	435	2	US-08-531-439B-4	Sequence 4, Appli
38	114	4.5	1128	4	US-08-923-992A-6	Sequence 6, Appli
39	112.5	4.4	467	2	US-08-686-599A-17	Sequence 17, Appl
40	110.5	4.4	1066	4	US-09-541-782-8	Sequence 8, Appli
41	110.5	4.4	1184	4	US-09-541-782-2	Sequence 2, Appli
42	110	4.3	976	4	US-09-104-324B-4	Sequence 4, Appli
43	108	4.3	2285	4	US-09-308-375-2	Sequence 2, Appli
44	107	4.2	1164	4	US-08-923-992A-10	Sequence 10, Appl
45	107	4.2	1497	1	US-08-623-679-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-268-992-2
; Sequence 2, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-2

Query Match	98.3%	Score 2487;	DB 4;	Length 495;
Best Local Similarity	99.8%	Pred. No. 2.6e-229;		
Matches 470;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
QY	7	SNSGNMKPPLLVFIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDADEVKKALTG	66	
Db	25	NNSGNMKPPLLVFIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDADEVKKALTG	84	
QY	67	IKOMKIMMERKEKHTNLMSTLKKCKREEKQKALKEVQHLKEEERLCRESLADSWGE	126	
Db	85	IKOMKIMMERKEKHTNLMSTLKKCKREEKQKALKEVQHLKEEERLCRESLADSWGE	144	
QY	127	CRSCLENNCHRIYTCQPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIBEDA	186	
Db	145	CRSCLENNCHRIYTCQPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIBEDA	204	
QY	187	OLTQMEDVFSQLTVDVNSLNRSENVFPMQOEFDQTFQSHFISDTDLTEPYFFPAFSKE	246	
Db	205	OLTQMEDVFSQLTVDVNSLNRSENVFPMQOEFDQTFQSHFISDTDLTEPYFFPAFSKE	264	
QY	247	PMTKADLEOCWDIPNFFQFCNFVSIVSVSETITKMLKAIEDLPKODKAPDHGGLISK	306	

Db 265 PMTKADLEQCDWIDPNFQFCNFSVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLISK 324
Qy 307 MLPQDGRGLCGELDQNLSCRFKFKHCKQKQAHLSDCDPDPALHTLDEAIRLVNVSQ 366
Db 325 MLPQDGRGLCGELDQNLSCRFKFKHCKQKQAHLSDCDPDPALHTLDEAIRLVNVSQ 384
Qy 367 QYQIQLQMTKRKLEDTAYLVKMRGQFGWVSELANQAPETEIFNSIQVVPRIHEGNTSK 426
Db 385 QYQIQLQMTKRKLEDTAYLVKMRGQFGWVSELANQAPETEIFNSIQVVPRIHEGNTSK 444
Qy 427 QDETMTDLSILPSSNFTLKIPLESAESSNFTIGYVAKALQHFKEHFKTW 477
Db 445 QDETMTDLSILPSSNFTLKIPLESAESSNFTIGYVAKALQHFKEHFKTW 495

RESULT 2
US-09-268-992-64
; Sequence 64, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-64

Query Match 97.4%; Score 2464; DB 4; Length 466;
Best Local Similarity 100.0%; Pred. No. 3.7e-227;
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 MKPPLLVIIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDADEEYKALTGKQMK 71
Db 1 MKPPLLVIIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDADEEYKALTGKQMK 60
Qy 72 IMMERKEKHTNLMSTLKKCKREKQKQALKLLNEVOEHLLEERLCRESLADSWGCRSL 131
Db 61 IMMERKEKHTNLMSTLKKCKREKQKQALKLLNEVOEHLLEERLCRESLADSWGCRSL 120
Qy 132 ENNCNRIYTCQPSWSSVKNKIERFRKIYQFLFFPHEDNEKDLPISEKLIBEDAQLTOM 191
Db 121 ENNCNRIYTCQPSWSSVKNKIERFRKIYQFLFFPHEDNEKDLPISEKLIBEDAQLTOM 180
Qy 192 EDVFSQLTVDVNSLNRSNFVRQMOQEPDQTFQSHFISDSDLTPYFPFPAFSKEPMTKA 251
Db 181 EDVFSQLTVDVNSLNRSNFVRQMOQEPDQTFQSHFISDSDLTPYFPFPAFSKEPMTKA 240
Qy 252 DLEQCDWIDPNFQFCNFSVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPCQ 311
Db 241 DLEQCDWIDPNFQFCNFSVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPCQ 300
Qy 312 DRGLCGELDQNLSCRFKFKHCKQKQAHLSDCDPDPALHTLDEAIRLVNVSNOYQGOI 371
Db 301 DRGLCGELDQNLSCRFKFKHCKQKQAHLSDCDPDPALHTLDEAIRLVNVSNOYQGOI 360

Qy 372 LQMTKRKLEDTAYLVKMRGQFGWVSELANQAPETEIFNSIQVVPRIHEGNTSKODETM 431
Db 361 LQMTKRKLEDTAYLVKMRGQFGWVSELANQAPETEIFNSIQVVPRIHEGNTSKODETM 420
Qy 432 MTDLSTLPSSNFTLKIPLESAESSNFTIGYVAKALQHFKEHFKTW 477
Db 421 MTDLSTLPSSNFTLKIPLESAESSNFTIGYVAKALQHFKEHFKTW 466

RESULT 3
US-09-268-992-4
; Sequence 4, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-4

Query Match 94.0%; Score 2377; DB 4; Length 477;
Best Local Similarity 91.5%; Pred. No. 8e-219;
Matches 455; Conservative 1; Mismatches 1; Indels 40; Gaps 2;

Qy 1 MRETWYSNGNKKPPLLVFIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDADEV 60
Db 1 MRETWYSNGNKKPPLLVFIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDADEV 60
Qy 61 KKALTGIGKQKIMMERKER-----EHTNLMSTLKKCKREKQKQALK 100
Db 61 KKALTGIGKQKIMMERKERKANQAPETEIFNSIQVVPRIEHTNLMSTLKKCKREKQKQALK 120
Qy 101 LLNEVOEHLLEERLCRESLADSWGCRSCLNENCMRIYTCQPSWSSVKNKIERFRKI 160
Db 121 LLNEVOEHLLEERLCRESLADSWGCRSCLNENCMRIYTCQPSWSSVKNKIERFRKI 180
Qy 161 YOFLPFPHEDNEKDLPISEKLIBEDAQLTOMEDVFSQLTVDVNSLNRSNFVRQMOQEF 220
Db 181 YOFLPFPHEDNEKDLPISEKLIBEDAQLTOMEDVFSQLTVDVNSLNRSNFVRQMOQEF 240
Qy 221 DQTFQSHFISDSDLTPYFPFPAFSKEPMTKADLEQCDWIDPNFQFCNFSVSIYESVSET 280
Db 241 DQTFQSHFISDSDLTPYFPFPAFSKEPMTKADLEQCDWIDPNFQFCNFSVSIYESVSET 300
Qy 281 ITKMLKAIEDLPKQDKAPDHGGLISKMLPGQDRGLCGELDQNLSCRFKFKHCKQKQAHL 340
Db 301 ITKMLKAIEDLPKQDKAPDHGGLISKMLPGQDRGLCGELDQNLSCRFKFKHCKQKQAHL 360
Qy 341 SEDCPDVPALHTLDEAIRLVNVSNOYQGOIQLQMTKRKLEDTAYLVKMRGQFGWVSEL 400
Db 361 SEDCPDVPALHTLDEAIRLVNVSNOYQGOIQLQMTKRKLEDTAYLVKMRGQFGWVSEL - 419
Qy 401 NOAPETEIFNSIQVVPRIHEGNTSKODETMTDLSILPSSNFTLKIPLESAESSNFTIG 460
Db 420 -----HEGNTSKODETMTDLSILPSSNFTLKIPLESAESSNFTIG 460

QY 461 YWAKALQHKEHFKTW 477
 Db 461 YWAKALQHKEHFKTW 477

RESULT 4

US-09-268-992-51
 : Sequence 51, Application US/09268992
 : Patent No. 6342351
 : GENERAL INFORMATION:
 : APPLICANT: Chen, H.
 : APPLICANT: Freimer, N.
 : TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
 : FILE REFERENCE: 7853-138
 : CURRENT APPLICATION NUMBER: US/09/268,992
 : EARLIER FILING DATE: 1999-03-16
 : EARLIER APPLICATION NUMBER: 09/236,134
 : EARLIER FILING DATE: 1999-01-22
 : EARLIER APPLICATION NUMBER: 60/106,056
 : EARLIER FILING DATE: 1998-10-28
 : EARLIER APPLICATION NUMBER: 60/088,312
 : EARLIER FILING DATE: 1998-06-05
 : EARLIER APPLICATION NUMBER: 60/078,044
 : EARLIER FILING DATE: 1998-03-16
 : NUMBER OF SEQ ID NOS: 84
 : SOFTWARE: FastSeq for Windows Version 3.0
 : SEQ ID NO 51
 : LENGTH: 446
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 US-09-268-992-51

Query Match 92.9%; Score 2350; DB 4; Length 446;
 Best Local Similarity 100.0%; Pred. No. 2.8e-216;
 Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 APTWKDKTATSENLSKSEVGEIDAEVKKALTGIGKIMMERKEHTNLMSTLKK 91
 Db 1 APTWKDKTATSENLSKSEVGEIDAEVKKALTGIGKIMMERKEHTNLMSTLKK 60

QY 92 REEQEALKLLNEVQHEERLCRESLADSWGECRSCLENNCMRYTTCQPSWSSVKN 151
 Db 61 REEQEALKLLNEVQHEERLCRESLADSWGECRSCLENNCMRYTTCQPSWSSVKN 120

QY 152 KIERFRKIYQFLPPFHEDNEKDLPISEKLIEDAQLTOMEDVFSQLTVDVNSLFRSFN 211
 Db 121 KIERFRKIYQFLPPFHEDNEKDLPISEKLIEDAQLTOMEDVFSQLTVDVNSLFRSFN 180

QY 212 VFQMQQEFDTQSFHISDTLTERPYFFPAFSKEPMTKADLEQCDIPNFFQFCNFSV 271
 Db 181 VFQMQQEFDTQSFHISDTLTERPYFFPAFSKEPMTKADLEQCDIPNFFQFCNFSV 240

QY 272 SIYESVETITKMLKATEDLPKQKADPHGGLISKMLPGQDGLCGELDONLSRCFPHE 331
 Db 241 SIYESVETITKMLKATEDLPKQKADPHGGLISKMLPGQDGLCGELDONLSRCFPHE 300

QY 332 KCQKQAHLSQDPCDVPALHTDELALRLVNSVNOQYQGLIQMTKRHLEDATYILVEKMRG 391
 Db 301 KCQKQAHLSQDPCDVPALHTDELALRLVNSVNOQYQGLIQMTKRHLEDATYILVEKMRG 360

QY 392 QFGWVSELANOAPTEIIFNSIQVVPRIHEGNTSKQDETMTDLSILPSSNFTIKIPLLEE 451
 Db 361 QFGWVSELANOAPTEIIFNSIQVVPRIHEGNTSKQDETMTDLSILPSSNFTIKIPLLEE 420

QY 452 SAESSNFIGYVWAKALQHKEHFKTW 477
 Db 421 SAESSNFIGYVWAKALQHKEHFKTW 446

RESULT 5

US-09-268-992-49
 : Sequence 49, Application US/09268992
 : Patent No. 6342351
 : GENERAL INFORMATION:
 : APPLICANT: Chen, H.
 : APPLICANT: Freimer, N.
 : TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
 : FILE REFERENCE: 7853-138
 : CURRENT APPLICATION NUMBER: US/09/268,992
 : EARLIER FILING DATE: 1999-03-16
 : EARLIER APPLICATION NUMBER: 09/236,134
 : EARLIER FILING DATE: 1999-01-22
 : EARLIER APPLICATION NUMBER: 60/106,056
 : EARLIER FILING DATE: 1998-10-28
 : EARLIER APPLICATION NUMBER: 60/088,312
 : EARLIER FILING DATE: 1998-06-05
 : EARLIER APPLICATION NUMBER: 60/078,044
 : EARLIER FILING DATE: 1998-03-16
 : NUMBER OF SEQ ID NOS: 84
 : SOFTWARE: FastSeq for Windows Version 3.0
 : SEQ ID NO 49
 : LENGTH: 465
 : TYPE: PRT
 : ORGANISM: Bos sp.
 US-09-268-992-49

Query Match 69.4%; Score 1755.5; DB 4; Length 465;
 Best Local Similarity 70.8%; Pred. No. 1.9e-159;
 Matches 330; Conservative 58; Mismatches 77; Indels 1; Gaps 1;

QY 12 MKPPLVFIYVLLWLDKSHCAPTWKDKTATSENLSKSEVGEIDAEVKKALTGIGKIM 71
 Db 1 MKPPLVFIYVLLWLDKSHCAPTWKDKTATSENLSKSEVGEIDAEVKKALTGIGKIM 60

QY 72 IMMERKEHTNLMSTLKKREEKQEAALKLLNEVQHEERLCRESLADSWGECRSC 131
 Db 61 ILMERREEHSLMRTLLKCKREEKQEAALKLLNEVQHEERLCRESLADSWGECRSC 120

QY 132 ENNCMYTTCQPSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLIEDAQLTOM 191
 Db 121 ESDCMREYTTCCQSWSSMKSTIERFRKIYQFLPPFHEDNEKDLPISEKLIEDAQLTOM 180

QY 192 EDVFSQLTVDVNSLFRSFNVEROMQOEFDQTFQSHFISDTLTERPYFFPAFSKEPMTKA 251
 Db 181 ENVFSQLTVDVNSLFRSFNVEROMQOEFDQTFQSHFISDTLTERPYFFPAFSKEPMTKA 240

QY 252 DLQCDWDPNFFQFCNFSVSIYESVETITKMLKATEDLPKQKADPHGGLISKMLPGQ 311
 Db 241 HPMQSWDIPNFFQFCNFSVSIYESVETITKMLKATEDLPKQKADPHGGLISKMLPGQ 300

QY 312 DRGLCGELDONLSRCFPHEKQKQAHLSQDPCDVPALHTDELALRLVNSVNOQYQGI 371
 Db 301 GRGLCGEPGQNSSECLQFHARQKQCDYLWADCPAVPELYTKADEALELYNISNQYQAV 360

QY 372 LQMTKRHLEDATYILVEKMRGQFGWVSELANOAPTEIIFNSIQVVPRIHEGNTSKQDET 431
 Db 361 LQMTKRHLEDATYILVEKMRGQFGWVSELANOAPTEIIFNSIQVVPRIHEGNTSKQDET 420

QY 432 MTDLISILPSSNFTIKIPLLESAESSNFIGYVWAKALQHKEHFKTW 477
 Db 421 I-DISILPSSNFTIKIPLLESAESSNFIGYVWAKALQHKEHFKTW 465

RESULT 6

US-09-268-992-39
 : Sequence 39, Application US/09268992
 : Patent No. 6342351
 : GENERAL INFORMATION:
 : APPLICANT: Chen, H.
 : APPLICANT: Freimer, N.
 : TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING

; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Cavia sp.
US-09-268-992-39

Query Match 64.8%; Score 1639; DB 4; Length 466;
Best Local Similarity 66.7%; Pred. No. 2.7e-148;
Matches 312; Conservative 63; Mismatches 89; Indels 4; Gaps 3;
QY 12 MKPPLLVFIVCLLWKDCHCAPTWKDTAISENLKSFSEVGIEDADEEVKKALGTGKQMK 71
Db 1 MKPLPLMPVCLLWKDCHCAPTWKDTAISENANFSFSEAGEIDVDGEVKIALIGIKQMK 60
QY 72 IMMERKEEHTNLMSTLKKCKEEKQKALKLNVEOHLNEVEHLEERLCRESLADSWGCRSCL 131
Db 61 IMMERREEHSLKMLTKLKKCKEEKQKALKLNVEHLEHLEESLQCVSLADSWDECRACL 120
QY 132 ENNCRIYITTCOPSSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIIEEDAQLTQM 191
Db 121 ENSCMRFDITTCOPAWSSVKNMVEQFRKIYQFLFPLQE--NDRSGPVSKGVTEEDAQVSHI 179
QY 192 EDVFSOLTVDVNSLNRNFRNQOEEDQTFQSHFISDTDLTPEYFPFAPSKPEMTKA 251
Db 180 EHVSQLSADVTSLNRSLYVFKLRREFDQAFQSYFTSGTDVTEFFFPPLSKPEPAYRA 239
QY 252 DLEOCWDIPNFQFCNFSVSVESYSETITKMLKAIEDLPKQDAPDHGGLISKMLPGQ 311
Db 240 DAEPWAIINPVQLLNLFSVYQSVSEKLIITLRATEDPPKQDKDSNOGGPISKILPEQ 299
QY 312 DRGLGELDONLSRCFKFHEKQKQAHLSDECDPVPALHTELDEAIRLVNVSNOQYGOI 371
Db 300 DRGSDGKLGONLSDCVNFRKRCQKQDYLSDDCPNVPPELYRELNEALRLVSRNQYDQV 359
QY 372 LQWTRKHLDETAYLVKEMRGQFGWVSELANOAPETELIENSIOVVPRI--HEGNISKODE 429
Db 360 VQMTQYHLEDITLLMEKMRQFGWVSELAYQSPGAEIDFPNVKVMVALSAHEGNSDQDD 419
QY 430 TMTDLSILPSSNFTLKIPLESASNSNFIQYVVAKALQHFKHEFKTW 477
Db 420 TVVPS-SLLPSSNFTLSSPLEKSAGNANFDHVVKEVLQHFKEFKTW 466

RESULT 7
US-09-268-992-41
; Sequence 41, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056

; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Cavia sp.
US-09-268-992-41

Query Match 61.6%; Score 1558; DB 4; Length 450;
Best Local Similarity 63.9%; Pred. No. 1.4e-140;
Matches 299; Conservative 62; Mismatches 87; Indels 20; Gaps 3;
QY 12 MKPPLLVFIVCLLWKDCHCAPTWKDTAISENLKSFSEVGIEDADEEVKKALGTGKQMK 71
Db 1 MKPLPLMPVCLLWKDCHCAPTWKDTAISENANFSFSEAGEIDVDGEVKIALIGIKQMK 60
QY 72 IMMERKEEHTNLMSTLKKCKEEKQKALKLNVEOHLNEVEHLEERLCRESLADSWGCRSCL 131
Db 61 IMMERREEHSLKMLTKLKKCKEEKQKALKLNVEHLEHLEESLQCVSLADSWDECRACL 120
QY 132 ENNCRIYITTCOPSSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIIEEDAQLTQM 191
Db 121 ENSCMRFDITTCOPAWSSVKNM-----ENDRSRGPVSKGVTEEDAQVSHI 163
QY 192 EDVFSOLTVDVNSLNRNFRNQOEEDQTFQSHFISDTDLTPEYFPFAPSKPEMTKA 251
Db 164 EHVSQLSADVTSLNRSLYVFKLRREFDQAFQSYFTSGTDVTEFFFPPLSKPEPAYRA 223
QY 252 DLEOCWDIPNFQFCNFSVSVESYSETITKMLKAIEDLPKQDAPDHGGLISKMLPGQ 311
Db 224 DAEPWAIINPVQLLNLFSVYQSVSEKLIITLRATEDPPKQDKDSNOGGPISKILPEQ 283
QY 312 DRGLGELDONLSRCFKFHEKQKQAHLSDECDPVPALHTELDEAIRLVNVSNOQYGOI 371
Db 284 DRGSDGKLGONLSDCVNFRKRCQKQDYLSDDCPNVPPELYRELNEALRLVSRNQYDQV 343
QY 372 LQWTRKHLDETAYLVKEMRGQFGWVSELANOAPETELIENSIOVVPRI--HEGNISKODE 429
Db 344 VQMTQYHLEDITLLMEKMRQFGWVSELAYQSPGAEIDFPNVKVMVALSAHEGNSDQDD 403
QY 430 TMTDLSILPSSNFTLKIPLESASNSNFIQYVVAKALQHFKHEFKTW 477
Db 404 TVVPS-SLLPSSNFTLSSPLEKSAGNANFDHVVKEVLQHFKEFKTW 450

RESULT 8
US-09-268-992-43
; Sequence 43, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0


```

; SEQ ID NO 43
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Cavia sp.
US-09-268-992-43

```

Query Match 50.5%; Score 1277; DB 4; Length 374;
Best Local Similarity 54.5%; Pred. No. 7.5e-114;
Matches 255; Conservative 44; Mismatches 73; Indels 9

Qy	12	MKPPLLVFI	VCLLWLK	DSCAPTWK	OKTATSE	NKLSFSE	VEIDAEE	VKKALT	GIQKMK	71																		
Db	1	1	1	1	1	1	1	1	1	1																		
Qy	72	IMMERKE	HTNLMS	TLKKCKE	QKAL	LLNVSQ	HEERL	CRSLAD	SGWGRSCL	131																		
Db	61	IMMERRE	EEHSK	LMKTL	KKCKE	QKAL	LMNVEH	EEESL	COVSLAD	SWDECRACL	120																	
Qy	132	ENCMRI	YTTCP	SWSSW	NKNK	IERFR	KIYQ	FLFP	PHEDN	EXDLP	ISEKLI	EEDAQ	LTO	M	191													
Db	121	ESNCMR	EDTT	TCQ	PAW	SSVK	----	----	----	----	----	----	----	----	140													
Qy	192	EDVFSQ	LTVD	NSL	FNRS	FN	VFMQ	QEF	DTQ	FQSH	FISD	TLTE	PP	FP	PA	SK	EP	MT	KA	251								
Db	141	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	147								
Qy	252	DLEQC	WDIP	NFFOL	FNFS	YI	SV	YSE	ITK	ML	KA	IEDL	PKQ	KAD	PHD	HGGL	IS	KML	PGQ	311								
Db	148	DABPS	WAIP	NV	FWL	LCNL	SP	VQ	SV	QSE	KL	TL	TR	ATED	PP	KQ	DS	NQ	GG	P	SK	IL	PI	BQ	207			
Qy	312	DRGLCE	LQNL	SR	CFE	KKCQ	KQ	QA	HL	SE	DC	PV	PA	LH	TE	LD	E	AI	R	LV	N	V	S	NQ	YQ	371		
Db	208	DRGSD	GKLG	QNL	SDC	VN	FR	KR	CQ	KQ	QD	Y	LS	DD	C	P	N	P	E	L	Y	R	E	L	N	A	371	
Qy	372	LQMT	RK	HE	LD	TAY	L	V	E	K	M	R	G	Q	F	C	W	S	E	L	A	N	A	P	E	T	E	429
Db	268	VQMTQ	VH	LE	DT	LL	M	E	K	M	R	E	Q	F	C	W	S	E	L	A	N	A	P	E	T	E	I	429
Qy	430	TNMTD	SIL	PSS	NFT	L	K	I	P	L	E	S	A	E	S	N	F	I	G	Y	V	A	K	A	L	Q	H	477
Db	328	TVVPS	-SLL	PSS	NFT	L	S	S	P	L	E	K	S	A	G	N	A	N	F	I	D	H	V	E	K	L	Q	477

```

RESULT          9
US-09-268-992-45
; Sequence 45, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Cavia sp.
US-09-268-992-45

```

Query Match	50.3%	Score 1272.5;	DB 4;	Length 373;
Best Local Similarity	54.5%;	Pred. No. 2e-113;		
Matches 255: Conservative	45;	Mismatches 71;	Indels	97;

Qy	12	MKPELLVIVCLLWKDCHCAPTWKKOKTAJSENKSFSEYGEIDADEEVKKALTGKOMK	71
Dd	1	MKLPELPVPVCLLWKDCHCAPTWKKOKTAJSENANSPSEAGEIDVDGVKIALIGIKQM	60
Qy	72	IMMERKEKHETNLMSTLKCKREEKOALKLLNEVOEHLBEERLCRESLADSGECRSL	131
Dd	61	IMMRREEHSHKMLTKCKEEKOALKLMNEVHEHLEESICQVSGLADSWDECRACL	120
Qy	132	ENNCMRIYTTCQPSSWSVNKKIERFFRKIYQFLPPFHEDNEKDLPISEKLIBSDAOLTM	191
Dd	121	ESNCMRDTTCCQPAWSSVKN-----	140
Qy	192	EDVFSQLTVDVNSLFNRSPNVFQQMOQBFTQTQSHPISDTDLTEPYFPAPFSKEPMTKA	251
Dd	141	-----MPAY-----BA	146
Qy	252	DLEOCWDIPNPFQFCNFYSIVSESITKMLKAIEDLPKODKPADHGGLISKMLPGQ	311
Dd	147	DAEPSWAIPNVFOLLNLNLSFVYGSVEKLIITLRATEDPPKODKSNOGGPSKILPEQ	206
Qy	312	DRGLCGELDONLSRCFFEHEKCKQAHLESDECDPDVALHTLEDJAIRLVNYSNQYGOI	371
Dd	207	DRGSDGLQGNLSDCVNFRCKOCBDYLDSDCPNPVELYRELNEALRLVSRSNQOYDOV	266
Qy	372	LQMTRKHELDYATVLVEKMRQFGVWSLANQAPETEIFNISIOGVPRI--HEGNISKQDE	429
Dd	267	VQMTQYHLEDTTLLMEKMRQFGVWSLAYQSCEADIFNPVKVMALSNEGNSDDDD	326
Qy	430	TMWTDLSILPSSNTTLKIPIEESAESSNFTGYVVAKALOQHFKBHKTW	477
Dd	327	TWVPS-SLLPSSNTLSPLEKAGSAGNAFIDHWKVYLQHFKEGHFKTW	373

```

RESULT 10
US-09-268-992-67
: Sequence 67, Application US/09268992
: Patent No. 6342351
: GENERAL INFORMATION:
: APPLICANT: Chen, H.
: APPLICANT: Freilmer, N.
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
: TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
: FILE REFERENCE: 7853-138
: CURRENT APPLICATION NUMBER: US/09/268,992
: CURRENT FILING DATE: 1999-03-16
: EARLIER APPLICATION NUMBER: 09/236,134
: EARLIER FILING DATE: 1999-01-22
: EARLIER APPLICATION NUMBER: 60/106,056
: EARLIER FILING DATE: 1998-10-28
: EARLIER APPLICATION NUMBER: 60/088,312
: EARLIER FILING DATE: 1998-06-05
: EARLIER APPLICATION NUMBER: 60/078,044
: EARLIER FILING DATE: 1998-03-16
: NUMBER OF SEQ ID NOS: 84
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 67
: LENGTH: 208
: TYPE: prt
: ORGANISM: Homo sapiens
US-09-268-992-67

```

```

Query Match      30.7%; Score 776; DB 4; Length 208;
Best Local Similarity 98.6%; Pred. No. 2.4e-66;
Matches 145; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 7 SNSNGMKPPLLVFIVCLLWLKDSHCAPWVKOKTATSENLSKFSEYGEIDADEVKALTG 66
      :|||||
Db 57 NNSGNMKPPLLVFIVCLLWLKDSHCAPWVKOKTATSENLSKFSEYGEIDADEVKALTG 116
      :|||||

```

QY 67 IKQKIMMERKEHTNLMSTLKKCREKQKALKLLNEVQEHLEBEERLCRESLADSGE 126
Db 117 IKQKIMMERKEHTNLMSTLKKCREKQKALKLLNEVQEHLEBEERLCRESLADSGE 176
QY 127 CRSCLENNCMRIYTTCPQSWSSVKNKI 153
Db 177 CRSCLENNCMRIYTTCPQSWSSVKNKL 203
RESULT 11
US-08-685-576-1
; Sequence 1, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Koza
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-325129
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-685-576-1
Query Match 5.7%; Score 145; DB 2; Length 1388;
Best Local Similarity 16.5%; Pred. No. 8.6e-05;
Matches 94; Conservative 97; Mismatches 166; Indels 212; Gaps 20;
QY 36 KDKTAISENLKSFSEVGIDEADVEYKALTGKQMKI-----MMERKEKHTNLM 86
Db 780 KQKVLNEDVRNLT---LKIQETQKRLTQNDLKMQTQVNTLMKSEKQLKQENHLL 835
QY 87 TLK-----KCREKQKALKLLNEVQEHLEEE-----RLCR 117

Db 836 EMKMSLERQNAELRRKQDADQCMKELQDLAEQYFSTLYKTQVRELKECECKTKLCK 895
QY 118 E-----SLADSWGCRSCLENNCMRIYTTCPQSWSSVKNKIERFERFIYQFLPPFHEDN 171
Db 896 ELQKKQELQDE-----RSLAAQLEITLLTKADSEQLARIAEEQYSDLEKE 942
QY 172 E--KDLPISE-----KLIEEDAQLTQMEDVFSQLTVDVNSLFNRSFNVFRMQQBFQDQ 222
Db 943 KIMLEIKEMMARHKQELTEKDATIASLEETNRTLSDVANLANEKEELNNKLEAQEQ 1002
QY 223 TFQSHFISDTLTPEYFFPAFSKEPMTKADLEQCWDIPNFOLFQCNFSVSYESVETIT 282
Db 1003 LSR---LKDEEISAAAIKAQFEKQLLTERTLK-----TOAVN 1036
QY 283 KMLKAIE-----DLPKQDKAPDHGGLISKMLPGQDRGLGEL-----DQNLRCF 327
Db 1037 KLAELMNRKEPVKRGNDTVRRKE-----ENKKLHMLKSEKKTQOMI 1082
QY 328 KFHEKQKQAHLSDDCPDPALHTEDELAIRLVNSQQYQGILOMTRKHLE-----380
Db 1083 KYQKELNEMQAIAE---SQIRIELQMTLDSKSDIEQLRSQALHIGLIDSSSIGSG 1138
QY 381 -----DTAYLVEKMRG-----QFGWSEL-----399
Db 1139 PGDEADGFPESRLEGWLSLFPVRNNTKKGWVKYIVSVSSKKILFYDSEQDKQSNPYM 1198
QY 400 -----ANQAPETIIFNSIQVVPRI-----HEGNISKQDETMMDLSILPSSNFT 444
Db 1199 VLDIDKLFHVRPVTQDVYRADAKIPIRILYLANEGESKKEQ-----1243
QY 445 LKIPLESAESSENF---GYVYAKALQHF 470
Db 1244 --FPVEPVGEKSNYICHKGHEIPTLYHF 1270
RESULT 12
US-08-685-576-4
; Sequence 4, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Koza
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-325129
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996

ATTORNEY/AGENT INFORMATION:
 NAME: Bent, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 16887/843
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1388 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-685-576-4

Query Match 5.68; Score 141; DB 2; Length 1388;
 Best Local Similarity 16.88; Pred. No. 0.00021;
 Matches 94; Conservative 97; Mismatches 176; Indels 192; Gaps 20;

36 KDKTAISENLKSFSEVGEIDAEEVKKALTGKIKOMKI-----MMERKEKEHTNLS 86
 780 KQDVLENDVRLNLA---LKIEQETQKRLTQNDLKQWQVNTLKMSEKQLKQENHLM 835
 87 TLK-----KREKQEAALKLLNEVOHLEEEE---RLCRESLADSWGEC----- 127
 836 EMKNLEKQNAELKREKQADQGMKELQDLQAEQYFSTLYKTQVRELKECEKTKLKG 895
 128 ---RSCLENNCMRIYTCQPSWSSVKKKIERFFRKIYQFLPPFHEDNE--KDLPISE--- 179
 896 ELQKKQLQELQDERSLAQLAETITKADSEQLARSIAEEOYSLEKEKIMKEIKEMMA 955
 180 ----KLIEEDAQLQMEDVFSQLTVDVNSLNFNRS---FNVFRMQOEFDOTFQSHFISDT 232
 956 RHKQELTEKQATIASLEETNRLTSDVANLANEKEELNKKLDVQEOQLSR-----LKDE 1009
 233 DLTPYFPFAPSKPMYKADLEQWCWDIPNFFQLFCNFSVSIYESVETITKMLKAIE--- 289
 1010 EISAAAIKAQPEKOLLERTLTK-----TQAVNKLAEIMNRKE 1046
 290 -----DLPKQDKAPDHGGLSKMLPGQDRGLCGEL---DNLSRCFKFHEKCKQCK 337
 1047 PVKRGNTDVRKEK-----ENRKLHMLKEREKLTQOMIKYOKELNEMQ 1092
 338 AHLSEDCDVPALHTELDEAIRLVNSNQYGOILQMTKRHLE-----DTAY 384
 1093 AQIAEE---SQIRLEQMTLSDKSDIEQLRSQALHIGLSDSSSTGSGPGAADGGF 1148
 385 LVERKRG-----QFGWYSEL-----A 400
 1149 PESRLEGWLSLVVRNNTKFGWKYVIVSVSKILFYDSEQDKBQSNPYMVLIDIKLFHV 1208
 401 NQAPETEIFIENSIOVVPRI-----HSGNISKQDETMTDLSILPSSNFTLKIPLES 454
 1209 RPVTQTDVYRADAKEIRIPIQLYANEGESKKEOE-----FPVEPVGE 1251
 455 SSNFI---GYVAKALQHF 470
 1252 KSNYICHKGHFIPTLYHF 1270

RESULT 13
 US-08-328-254-6
 Sequence 6, Application US/08328254
 Patent No. 5710022
 GENERAL INFORMATION:
 APPLICANT: Zhu, Xuelliang
 TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700
 CITY: San Diego
 STATE: California
 COUNTRY: USA
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/328,254
 FILING DATE: 24-OCT-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/141,239
 FILING DATE: 22-OCT-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-CJ 1191
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2482 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-328-254-6

Query Match 5.28; Score 130.5; DB 1; Length 2482;
 Best Local Similarity 18.88; Pred. No. 0.0051;
 Matches 99; Conservative 85; Mismatches 165; Indels 177; Gaps 22;

27 KDSHCAPTMDKDTAISNLK-----SFSEVGE-----IDADE---EVKKALTGKIQ 69
 1502 KQLHIAEKLERENDSLADKVENLERELQMSSENOELVILDRENSKAETVTKQIEE 1561
 70 M-----KIMMERKEKEH-----TNLMSTLKKCREKQEAALKLLNEV 105
 1562 MARSKLVFELDLVTLRSEKENLTKQIQKQGLSELKLLSSFKSLEKEQA-----EI 1616
 106 QEHLSEERCLRESLADSWGECRSCLENNC--MRIYTCQPSWS-----SVKNKIERF 156
 1617 Q--IKESKTAVENTMQQLKELNEAVALCGDQOEIMKATEQSLDPPIEEHQLRNSIEKL 1674
 157 FRKIYQFLFPFHEDNEKDLPISSEKLIBED----- 185
 1675 RARL-----EADKKQLCVLQQLKESEHHDLLKGRVENLERELIARTNOEHALEA 1727
 186 -----AQLTQMEDVFSQLTVDVNSLNFNRSPVFRMQOEFDOTFQSHFISDTDL 234
 1728 ENSKGEVETLTKAKIEGTQSLRGLDELVDVTVIRSEKENTLNELOKEQER-----ISELEI 1781
 235 TEPYFFAPFSKPEMTKADLEQWCWDIPNFFQLFCNFSVSIYESVETITKMLKA--IEDLPK 293
 1782 INSSFENILOEKEQEKVQMAE-----KSSSTAMEMLOTQLKELNE 1820
 294 QDKAPDHGGLSKMLPGQDRGLCGELDQNLIS---RCFKFHEKCKQCAHLSDECDPVPAL 350
 1821 RVAALHN-----DQEAACKAEQNLSSQVECELEL-EKAQLLOG----- 1856
 351 HTLDEDAIRLVNSNQYGOILQMTKRHLEDYALVEKMRGQFGWGSSELANQAPETEIF 410
 1857 ---LDEAKNNYIVLQSSVNGLIQ-----EVEDGKQKLEKDEE---ISRLKNQIQDQEQVL 1906
 411 NSIQVVPRIH-----EGNISKQDETMTD--LSILPSSNFTLKIPLE 450
 1907 SKLSQVEGEHQLWKEQNLELRNLTVLEBQKIQVSKNASLODPTLE 1952

RESULT 14
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
US-08-353-700-1

Query Match 5.1%; Score 128.5; DB 1; Length 3248;
Best Local Similarity 18.8%; Pred. No. 0.012;
Matches 99; Conservative 85; Mismatches 165; Indels 177; Gaps 22;

QY 27 KUSHCAPTWKDKTAISENLK-----SFEVGE---IDADE---EVKALGTGIQ 69
DB 2230 KQLHIAEKLERENDSLKDKVENLERELQSEENQELVILDAENSKAEVETLKTQIEE 2289
QY 70 M-----KIMMERKEKEH-----TNLMSTLKKCKREKQKALKLLNEV 105
DB 2290 MARSLKIFDLVTLRSEKENLTKQIEKQOGQLSELDKLLSSPKSLLEEKEQA-----EI 2344
QY 106 QEHLBEERLCRESLADSGEGRSCLENNC--NRITTCQPSWS-----SVKNKIERF 156
DB 2345 Q---IKESKTAVEMLNQKQKELNEAALCGOQEIIMKATEQSLDPPPIEEHQLRNSIEKL 2402
QY 157 FRKIYQFLFPFHEDNKKLPISKELEED-----SFEVGE---IDADE---EVKALGTGIQ 185
DB 2403 RARL-----EADKQLCVLQOLKESEHHDLLKGRVENLERELEIARTNOEHALEA 2455
QY 186 -----AQLTQMEDVFSQTVVNSLFRNSFNVRQMQQDFDQTFQSHFISDTDL 234
DB 2456 ENSKGEVETLKKAKIEGTMQSLRGLELDVVTIRSEKENLTNLEQKEQER-----ISELEI 2509
QY 235 TEPYFFPAPSKPEMTKADLEQCWDIYFNFFQFCNFSVSIYESVETITKMLKA-IEDLPK 293

DB 2510 INSSFENILQEKQEKVOMKE-----KSSAMEMLQTLQKELNE 2548
QY 294 QDKAPDHGGLISKMLPGQDRGLCGELDQNLNLS---RCFKFHEKCKCQCAHLSDCPDVPAL 350
DB 2549 RVAALHN-----DOEACKAKEQNLSSQVECLEL-EKAQLLQG----- 2584
QY 351 HTEDEAIRLVNVSNOQYQGIQLQMTKRLHLEDTAYLVKMRGQFGWVSELANQAPETEIIIF 410
DB 2585 ----LDEAKNNYIVLQSSVKGLIQ-----EVEDGKQKLEKKDEE---ISRLKNOIQDQEQIV 2634
QY 411 NSIQVVPRIH-----EGNISKQDETMKTD---LSILPSSNFTLKIPL 450
DB 2635 SKLSQVEGHEQHWKQNLRLNLTVELQKIQVLOSKNASLQDTTLE 2680

RESULT 15
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 5.1%; Score 128.5; DB 5; Length 3248;
Best Local Similarity 18.8%; Pred. No. 0.012;
Matches 99; Conservative 85; Mismatches 165; Indels 177; Gaps 22;

QY 27 KUSHCAPTWKDKTAISENLK-----SFEVGE---IDADE---EVKALGTGIQ 69
DB 2230 KQLHIAEKLERENDSLKDKVENLERELQSEENQELVILDAENSKAEVETLKTQIEE 2289
QY 70 M-----KIMMERKEKEH-----TNLMSTLKKCKREKQKALKLLNEV 105
DB 2290 MARSLKIFDLVTLRSEKENLTKQIEKQOGQLSELDKLLSSPKSLLEEKEQA-----EI 2344

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:44:34 ; Search time 32.87 Seconds
(without alignments)
1394.421 Million cell updates/sec

Title: US-09-722-544A-4

Perfect score: 2529

Sequence: 1 MRTWDYSNGNMKPLLVFI.....FTGYVAKALQHFKEHFKTW 477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	434	17.2	449	1 A41386	clusterin precursor
2	416	16.4	451	2 I50131	clusterin - quail
3	410.5	16.2	445	2 A40018	clusterin precursor
4	406	16.1	449	2 S07714	T64 protein precursor
5	403	15.9	448	2 A40714	clusterin precursor
6	402	15.9	448	2 I56335	clusterin precursor
7	400.5	15.8	439	2 A35744	apolipoprotein J -
8	400.5	15.8	446	2 A42108	clusterin precursor
9	380	15.0	447	2 A27205	clusterin precursor
10	168	6.6	191	2 I48174	sulfated glycoprot
11	160	6.3	1738	2 T14867	interaptin - slime
12	151.5	6.0	1091	2 T34107	hypothetical prote
13	151	6.0	1526	2 T41522	myosin ii - fissio
14	146	5.8	1156	2 B70356	chromosome assembl
15	145.5	5.8	594	1 A44073	CIK1 protein - yea
16	145	5.7	1388	2 S70633	serine/threonine-s
17	144	5.7	3225	2 I52300	giantin - human
18	144	5.7	3259	1 A56539	giantin - human
19	142.5	5.6	2663	1 S28261	centromere protein
20	141.5	5.6	3433	1 S29381	utrophin - human
21	141	5.6	1005	2 A64465	hypothetical prote
22	140.5	5.6	2166	2 G70163	hypothetical prote
23	139	5.5	1955	2 T30934	myosin-like protei
24	137	5.4	1084	2 G13329	hypothetical prote
25	137	5.4	1300	2 I53799	CGI protein - huma
26	137	5.4	1940	1 S04090	myosin heavy chain
27	136	5.4	1132	2 T00259	hypothetical prote
28	135	5.3	1060	1 A40264	kinesin-related pr
29	134.5	5.3	1957	2 T38077	hypothetical colle

ALIGNMENTS

RESULT 1

A41386
clusterin precursor [validated] - human
N:Alternate names: apolipoprotein J; complement cytolysis inhibitor SP-40; complement TRPM-2/clusterin protein
C:Contains: clusterin alpha chain; clusterin beta chain
C:Species: Homo sapiens (man)
C:Date: 03-Apr-1992 #sequence_revision 17-Nov-1995 #text_change 08-Dec-2000
C:Accession: S43646; S04662; A41386; A35833; S34056; A53177; A37816; B37816; PL0136; R:Wong, P.; Taillefer, D.; Lakins, J.; Pineault, J.; Chader, G.; Tenniswood, M.
Eur. J. Biochem. 221, 917-925, 1994
A:Title: Molecular characterization of human TRPM-2/clusterin, a gene associated with A:Reference number: S43646; MUID:94237156
A:Accession: S43646
A:Molecule type: DNA
A:Residues: 1-449 <WON>
A:Cross-references: GB:M64722; NID:g339972; PIDN:AA806508.1; PID:g339973 R:Kirsbaum, L.; Sharpe, J.A.; Murphy, B.; d'Apice, A.J.F.; Classon, B.; Hudson, P.; EMBO J. 8, 711-718, 1989
A:Title: Molecular cloning and characterization of the novel, human complement-associ A:Reference number: S04662; MUID:89251601
A:Accession: S04662
A:Molecule type: mRNA
A:Residues: 1-449 <KIR>
A:Cross-references: EMBL:X14723; NID:g30250; PIDN:CAA32847.1; PID:g30251
A:Note: parts of this sequence, including the amino end of the mature protein, were c R:Jenne, D.E.; Tschopp, J.
Proc. Natl. Acad. Sci. U.S.A. 86, 7123-7127, 1989
A:Title: Molecular structure and functional characterization of a human complement cy tis fluid.
A:Reference number: A41386; MUID:89386692
A:Accession: A41386
A:Molecule type: mRNA
A:Residues: 2-449 <JEN>
A:Cross-references: GB:M25915; NID:g180619; PIDN:AAA35692.1; PID:g180620 R:de Silva, H.V.; Harmony, J.A.K.; Stuart, W.D.; Gil, C.M.; Robbins, J.
Biochemistry 29, 5380-5389, 1990
A:Title: Apolipoprotein J: structure and tissue distribution.
A:Reference number: A35833; MUID:90344779
A:Accession: A35833
A:Molecule type: mRNA
A:Residues: 34-449 <DES>
A:Cross-references: GB:J02908; NID:g178854; PIDN:AAA51765.1; PID:g178855 R:Ghiso, J.; Matsubara, F.; Koudinov, A.; Choi-Miura, N.H.; Tomita, M.; Wisniewski, T Biochem. J. 293, 27-30, 1993
A:Title: Research Communication. The cerebrospinal-fluid soluble form of Alzheimer's A:Reference number: S34056; MUID:93319521
A:Accession: S34056
A:Molecule type: protein
A:Residues: 228-240, 'X', 242-246; 23-24, 'X', 26-34, 'X', 36-38, 'X', 40-41 <GHI> R:James, R.W.; Hochstrasser, A.C.; Borghini, I.; Martin, B.; Pometta, D.; Hochstrasse Arterioscler. Thromb. 11, 645-652, 1991

kinectin 1 - human
myosin heavy chain
hypothetical prote
conserved hypotet
hypothetical prote
hypothetical prote
serine/threonine-s
hypothetical prote
myosin heavy chain
hypothetical prote
hypothetical prote
myosin-v - chicken
myosin heavy chain
cardiac muscle fac
protein C43E11.3
myosin heavy chain

30 134 5.3 1356 2 S32763
31 133.5 5.3 1313 2 A48467
32 133.5 5.3 1956 2 T16416
33 132.5 5.2 852 2 D72230
34 132 5.2 1033 2 T42701
35 132 5.2 1133 2 T22976
36 132 5.2 1388 2 S74245
37 132 5.2 1413 2 T26467
38 132 5.2 1992 2 A47297
39 131.5 5.2 1679 2 S48385
40 131 5.2 746 2 T21277
41 130.5 5.2 1830 1 S19188
42 130.5 5.2 1940 1 A24922
43 130 5.1 1538 2 T29095
44 130 5.1 1590 2 B87754
45 130 5.1 1961 1 A61231


```
Db 203 -----TPPFGGFEAF-VPPVQVRVHL-----VPRRLRLSRELHPFFQHPM 240
Qy 268 NFSYIYESVSETITKMLKAIEDLPKODKADPH--GGLISKMLP-GODRGLCGELDONLS 324
Db 241 HGFRLFPQLEPMTOHMLDG-----CHGAWHEPLGGFATESRNFSDRMVCRIRNSA 294
Qy 325 RCFRFEKQRCQCAHLSEDC-----PDVPALHTEDEAIRLVNVSNOQYGOQLQWTRKHLE 380
Db 295 GCLMRDCEKCRELLAVDCSQTPVQSOLREQFEDALRAERTRRYDDLLSAFAEML 354
Qy 381 DTAYLVEKMRQCFGWSELANQAP-----ETELIFNSIQVVPRIHSGNISKQDETMTWD 434
Db 355 NTSSLLDQLNRFQGWVSRGLNLTQGNDFLQVTTVFESK---TPNLEDP--SAPADTQVT- 408
Qy 435 LSLPSSNFTLKIPLEESAESSNFIQYVAKALQHF 470
Db 409 VOLPDSPLSLTVPGDISWDDPRFMEIVPEQALQHY 444

RESULT 5
clusterin precursor - mouse
N:Alternate names: sulfated glycoprotein 2
N:Contents: clusterin alpha chain; clusterin beta chain
C:Species: Mus musculus (house mouse)
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 13-Aug-1999
C:Accession: A40714; JN0699
R:French, L.E.; Chonn, A.; Ducrest, D.; Baumann, B.; Belin, D.; Wohlwend, A.; Kiss, J.Z.
J. Cell Biol. 122, 1119-1130, 1993
A:Title: Murine clusterin: molecular cloning and mRNA localization of a gene associated
A:Reference number: A40714; MUID:93359508
A:Accession: A40714
A:Molecule type: mRNA
A:Residues: 1-448 <FR>
A:Cross-references: GB:L08235; MID:g192596; PIDN:AAA37422.1; PID:g192597
R:Lee, K.H.; Ji, Y.M.; Lim, H.M.; Lee, S.C.; You, K.H.
Biochem. Biophys. Res. Commun. 194, 1175-1180, 1993
A:Title: Molecular cloning and sequencing of sulfated glycoprotein-2 cDNA from testis of
A:Reference number: JN0699; MUID:93356785
A:Accession: JN0699
A:Molecule type: mRNA
A:Residues: 1-448 <LE>
A:Note: the authors translated the codon CGC for residue 66 as Ala
C:Comment: This protein is involved in many different cellular processes other than those
C:Superfamily: clusterin
C:Keywords: apoptosis; disulfide bond; glycoprotein; spermatogenesis
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-226/Domain: clusterin beta chain #status predicted <BCH>
F:227-448/Domain: clusterin alpha chain #status predicted <MAT>
F:227-448/Domain: clusterin alpha chain #status predicted <ACH>
F:102,144,290,327,353,373/Binding site: carbohydrate (Asn) #status predicted

Query Match 15.9%; Score 403; DB 2; Length 448;
Best Local Similarity 23.4%; Pred. No. 5e-18;
Matches 112; Conservative 109; Mismatches 196; Indels 62; Gaps 14;

Qy 16 LLVFIIVCLWLKDSHCAPTWKDKTAISENLKSFSEVEIDA-----DEVKKALTGIKQ 69
Db 3 ILLLCVALLI-----WDNGMVLGEQVSDNELQELSTQGSRYINKEIQNAVQGVKH 54

Qy 70 MKIMMERKEHTNLMSTLKCKREEKQEAALKLLNEVQEHLEBEERLCRESLADSGEGRS 129
Db 55 IKTLIEKTAERKSLNLSLEAKKEDALEDTROSEMKLKAPPEVCNMTMALWEECKP 114

Qy 130 CLNNCMRIYT--TCOPSSWSVANKIERFRKIYQFLFPFHEDNEKDLPISEKLIBEDAQL 188
Db 115 CLKHTCMKFIYARVCRSGSLVGQOLEEFLNQSSPFFYFWNGDR-----IDSLESDRQ 168

Qy 189 TO-----MEDVFSQLTVDVNSLNRSPVFRMQQEQEDQTFQSHFISDTOLTPEYFPF 241
Db 169 SQVLDAMQDSFARASGIITLFDQRF--FAR-----ELHDHYFS-----PIGPPH 214
```

```
Qy 242 ---AFSKPEMTKADLEOC-WDIPNFFQPCNFVSIVSEYSETITKMLKAIEDLPKODKA 297
Db 215 PHFLYPKSRLVRSILMSPSHYGPSPFHNMFQFFEMIHQAQAAMDVLHSPAFQFPD 272
Qy 298 PDHGGLLSKMLPGQ-DRGLCGELDONLSRCFKFHEKQKQCAHLSEDC-----PDVPALHT 352
Db 273 -----FLREGEDDRTVCKEIRNRSTGCLKMKGQCEKQCEILSVDCSTNNPAQANLRQ 324
Qy 353 ELDEAIRLVNVSNOQYGOQLQWTRKHLEDYALVEKMRGQFGWVSELANOAPETELIFNS 412
Db 325 ELNDSLQVAERLTQYKELLQSFQSKMLNTSSLLEQLNDQFNWVSQLANLTQGEDKYLR 384
Qy 413 IQVVPRIHSGNISKQDETMTDLISLPSSNFTLKIPLEESAESSNFIQYVAKALQHF 471
Db 385 VSTV-TTHSSDSEVPVSRVTEVVVKLFDSDPITVVLPEVSKDNPKFMDTVAERKALQEYR 442

RESULT 6
apolipoprotein J - mouse
C:Species: Mus sp. (mouse)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 13-Aug-1999
C:Accession: I56335
R:Jordan-Starck, T.C.; Lund, S.D.; Witte, D.P.; Atonow, B.J.; Ley, C.A.; Stuart, W.D.
J. Lipid Res. 35, 194-210, 1994
A:Title: Mouse apolipoprotein J: characterization of a gene implicated in atherosclerosis
A:Reference number: I56335; MUID:94223204
A:Accession: I56335
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-448 <RES>
A:Cross-references: GB:S70244; NID:g546504; PIDN:AAB30623.1; PID:g546505
C:Genetics:
C:Gene: ApoJ
C:Superfamily: clusterin

Query Match 15.9%; Score 402; DB 2; Length 448;
Best Local Similarity 24.1%; Pred. No. 5.8e-18;
Matches 115; Conservative 108; Mismatches 201; Indels 54; Gaps 15;

Qy 12 MKPPLLFIIVCLWLKDSHCAPTWKDKTAISENLKSFSEVEIDADEVKALTGIKQ 70
Db 1 MKILLCVALLI--DNGMV---LGEQVSDNELQELSTQGSRYINKEIQNAVQGVKH 55

Qy 71 KIMMERKEHTNLMSTLKCKREEKQEAALKLLNEVQEHLEBEERLCRESLADSGEGRS 130
Db 56 KTLIEKTAERKSLNLSLEAKKEDALEDTROSEMKLKAPPEVCNMTMALWEECKP 115

Qy 131 LENNCMRIYT--TCOPSSWSVANKIERFRKIYQFLFPFHEDNEKDLPISEKLIBEDAQL 189
Db 116 LKHTCMKFIYARVCRSGSLVGQOLEEFLNQSSPFFYFWNGDR-----IDSLESDRQ 169

Qy 190 Q-----MEDVFSQLTVDVNSLNRSPVFRMQQEQEDQTFQSHFISDTOLTPEYFPF 241
Db 170 QVLDAMQDSFARASGIITLFDQRF--FAR-----ELHDHYFS-----PIGPPH 215

Qy 242 ---AFSKPEMTKADLEOC-WDIPNFFQPCNFVSIVSEYSETITKMLKAIEDLPKQKAP 298
Db 216 HFLYPKSRLVRSILMSPSHYGPSPFHNMFQFFEMIHQAQAAMDVLHSPAFQFPD 272

Qy 299 DHGGLSKMLPGQ-DRGLCGELDONLSRCFKFHEKQKQCAHLSEDC-----PDVPALHTE 353
Db 273 -----FLREGEDDRTVCKEIRNRSTGCLKMKGQCEKQCEILSVDCSTNNPAQANLRQ 325

Qy 354 LDEAIRLVNVSNOQYGOQLQWTRKHLEDYALVEKMRGQFGWVSELANOAPETELIFNSI 413
Db 326 LNDLSQVAERLTQYKELLQSFQSKMLNTSSLLEQLNDQFNWVSQLANLTQGEDKYLR 385

Qy 414 QVVPRIHSGNISKQDETMTDLISLPSSNFTLKIPLEESAESSNFIQYVAKALQHF 471
Db 386 STV-TTHSSDSEVPVSRVTEVVVKLFDSDPITVVLPEVSKDNPKFMDTVAERKALQEYR 442
```

RESULT 7

A35744
clusterin precursor - bovine
N:Alternate names: complement cytolysis inhibitor; glycoprotein III
N:Contains: clusterin alpha chain; clusterin beta chain
C:Species: Bos primigenius taurus (cattle)
C:Date: 05-Oct-1990 #sequence_revision 05-Oct-1990 #text_change 13-Aug-1999
C:Accession: A35744
R:Palmer, D.J.; Christie, D.L.
J. Biol. Chem. 265, 6617-6623, 1990
A:Title: The primary structure of glycoprotein III from bovine adrenal medullary chromaffin granules
A:Reference number: A35744; MUID:90216681
A:Accession: A35744
A:Molecule type: mRNA
A:Residues: 1-439 <PAL>
A:Cross-references: GB:J05391; NID:gl63114; PIDN:AAA30554.1; PID:gl63115
A:Note: parts of this sequence, including the amino ends of the alpha and beta chains, were determined from complementary DNA
C:Superfamily: clusterin
C:Keywords: apoptosis; disulfide bond; extracellular protein; glycoprotein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-221/Domain: clusterin beta chain #status experimental <BCH>
F:20-221,222-439/Product: clusterin #status experimental <MAT>
F:222-439/Domain: clusterin alpha chain #status experimental <ACH>
F:80,97,139,283,320,346,366/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 15.8%; Score 400.5; DB 2; Length 439;
Best Local Similarity 24.5%; Pred. No. 7e-18;
Matches 114; Conservative 97; Mismatches 183; Indels 71; Gaps 15;

QY 34 TWDKTAISE-NLKSFSEVEIDADEVKALTGIKIMMERKEKEHTNLMSTLKCR 92
DB 13 SWEGWAISDEKQEMSTEGSKYVNEIKNALKEVQIKTQIBQTEERKLLLSLEAK 72
QY 93 EEKQKALLNVEQHELEERLCRESLADSGEGRSCLENMCRITY-TQPSWSSVKN 151
DB 73 KKEDALNDRSENKLLKASQGVNCTMTALWEECKPCLKQTCMKFYARVCRSGSLVGH 132
QY 152 KIRFRFKYQVFFPHEDNEKDLPISEKLEDAQLTQMEDVFSQITVDVNSLFRSN 211
DB 133 QLEEFNLQSSPPFFWINGDR-----IDSLMENDREQSHVMDV-----MEDSFTRAS 179
QY 212 VFQMQQEDQFQSHFI---SDTLTEPY-----FF---PAFSKEPTKADLEQC 256
DB 180 I-----NFDHVFQPP---YDWHQAQAQMDAHLQ-----RTPYHPTM-EFTENDRTVC 232
QY 257 WDTPNFQLFNFVSIVESVSTITKMLKAIEDLPKQDKPDHGLISKMLPGQDRLC 316
DB 233 ----NFDHVFQPP---YDWHQAQAQMDAHLQ-----RTPYHPTM-EFTENDRTVC 277
QY 317 GELDONLSRCFHEKQCKQAHLSDEC-----PDVPALHTEDEAIRLVNSNQOYQIL 372
DB 278 KEIRHNSTGLRMKQDQCEKCEILEVDCASNPQTQTLRQQLNASLQAEKFSRLYDQLL 337
QY 373 QMTRKHEDTAYLVKMRGQFGWSELANOAPETE-----IIFNSIQVVPRIHGNISK 426
DB 338 QSVQKMLNTSALLKQLNQFTWVSQLANTQSDQHYLOVFTVNSHNSDPSIPSG----- 393
QY 427 QDETMMTDLISLPSSNFTLKIPLESSESSNFIYGVAKALQHF 471
DB 394 ----LTKVIVKLFNSPITVTPQVSSPNFMENVAEKALQOYR 433

RESULT 8

A42108
clusterin precursor - pig
N:Alternate names: complement cytolysis inhibitor; CP40 protein
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
C:Accession: A42108; JC5535; PC4475
R:Diemer, V.; Hoyle, M.; Baglioni, C.; Millis, A.J.
J. Biol. Chem. 267, 5257-5264, 1992

A:Title: Expression of porcine complement cytolysis inhibitor mRNA in cultured aortic
A:Reference number: A42108; MUID:92184774
A:Accession: A42108
A:Molecule type: mRNA
A:Residues: 1-446 <DIE>
A:Cross-references: GB:M84639; NID:gl64408; PIDN:AAA31013.1; PID:gl64409
A:Experimental source: aortic smooth muscle cells
A:Note: sequence extracted from NCBI backbone (NCBIN:87354, NCBI:87356)
R:Ogawa, S.; Ishibashi, Y.; Sakamoto, Y.; Kitamura, K.; Kubo, M.; Sakai, T.; Inoue, K.
Biochem. Biophys. Res. Commun. 234, 712-718, 1997
A:Title: The glycoproteins that occur in the colloids of senescent porcine pituitary
A:Reference number: JC5535; MUID:97318844
A:Accession: JC5535
A:Molecule type: DNA
A:Residues: 1-446 <OGA>
A:Accession: PC4475
A:Molecule type: protein
A:Residues: 58-66;68-77;229-247;249-251;408-436 <OG2>
C:Superfamily: clusterin
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-227/Domain: clusterin beta chain #status predicted <BCH>
F:23-227,228-446/Product: clusterin #status predicted <MAT>
F:228-446/Domain: clusterin alpha chain #status predicted <ACH>

Query Match 15.8%; Score 400.5; DB 2; Length 446;
Best Local Similarity 24.4%; Pred. No. 7.1e-18;
Matches 118; Conservative 98; Mismatches 196; Indels 71; Gaps 16;

QY 16 LLVFIIVCLLWLKDSHCAPTWK-----DKTALSENKLSFSEVEIDADEVKALTGIK 68
DB 4 LLLLVGLLL-----TWGNPVPVIGDKAISDKLEQEMSTEGSKYVNEIKNALKEYK 54
QY 69 QMKIMMERKEHTNLMSTLKCRKQKALKEQVHLEERLCRESLADSGEGR 128
DB 55 QIKTLIEQNEERKSLLSLEAKKKEDALNDRTRTETKLAGSGLCNETMMALWEECK 114
QY 129 SCLNNCMRIYT-TQPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLEEDAQ 187
DB 115 PCLKQTCMKFYARVCRSGSLVGHOLEEFLNQSSPPFYFWINGDR-----IDSLMENDRQ 168
QY 188 LTQMEDVFSQITVDVNSLFRSNFVRQMQQEDQFQSHFIS-----DPTDLTEPY----- 238
DB 169 QSHVMDI-----MEDSFNRASNI-----MDLFDQDRFRNRPFDQTFSPGSSHR 214
QY 239 ---FFPAPSKPEPTKADLEQCQWDIPNFOLFNFVSIVESVSTITKMLKAIEDLPKQD 295
DB 215 GSLFFNPKSRFARNIMPPFLFTDL-NYHDMFQPP-----FDMIHOQAQMDAHLRIPYH- 268
QY 296 KAPDHGLISKMLP--GODRGLCGELDONLSRCFKFHEKQCKQAHLSDEC-----PDVPA 349
DB 269 -FPEAG-----VPNSNDRAVCKEIRINSTGCLRMKQDQCEKREILSVDCSASNSQMQ 321
QY 350 LHTEDEAIRLVNSNQOYQILQMTKRKHEDTAYLVKMRGQFGWSELANOAPETEII 409
DB 322 LRQELYTSLOMAEKPSKLYDQLLSQYQKMLNTSLLKQLNEQFQSVQLANLTQDNDRY 381
QY 410 ENSIQVVPRIHGNISKQDETMMTDLISLPSSNFTLKIPLESSESSNFIYGVAKALQ 469
DB 382 YLOQVTV-NSHGSDPSVPFSGLTQVVVKLFDSYPTIPIQEVN--DPKFMETVAEALQ 438
QY 470 FKE 472
DB 439 YRQ 441

RESULT 9

A27205
clusterin precursor - rat
N:Alternate names: SCP-2; sulfated glycoprotein 2; TRPM-2
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 13-Aug-1999

Matches	102;	Conservative	88;	Mismatches	171;	Indels	128;	Gaps	21;
---------	------	--------------	-----	------------	------	--------	------	------	-----

Qy	36	KDKTAISENL-----KSFSEVG	Idadev	KKALTG	IKOMKIMMER-----	76									
Db	667	KDQNTINEQLNQLSEKDEIKLSKQOODEKINNLLLEIKKDC	LIERINQO	LLLEN	726										
Qy	77	-----KEKEHTNL	MASTLUK	KREEKOE	AKLLN	VEOHL	EEER	114							
Db	727	IDLNSKYOOLLLEFENFKLNSKEKE-----NQLNEL	QSKQDER	FNQLND--	EKLEKEQ	779									
Qy	115	LCRESLADSWGCRSCLNNCMRIVTTTCQPSWSSW	KNKIERF	KIYQFL	FPFHED	NEKD	174								
Db	780	L--QSIED	EFNQYK-----	QOQLSSNSN	IDOOL	OSTI	IELSEL	KEQKELN	822						
Qy	175	LPISKE	LIEDAQLTQ	MDVFSQ	LFT-----	VDVNS	LFNRSF	NVFR	QMOC	QEFQ	TQFOSH	227			
Db	823	---DSKLKE	KEKQLQO	LOQEF	DQLNEK	NOKDHQ	DOLELEKQ---	LKQLOQ	EYDOL	NETN	876				
Qy	228	FISD	TDLEPYFF	PAFSKE	PTKADLE-----	QCWD	IPNFFQ	LF	CNFS	VI	YESV	277			
Db	877	OSIENQL	NAQNL-----	INKEN	LEKEQEL	LKLQNL	NOQIEK	IQFDQ	QEF	SQNS	INIELV	933			
Qy	278	SETITKML	KAIED-----	LPQDK	DA---PDH	GLISK	MLPGD	RG	LCGL	EDQNL	SLRC	FPHE	331		
Db	934	NEKNEK	LIQLQD	YDQLKQ	NRSDER	KDEND	LIEK---ENQLK	STONEL	NOLIEK	NESD	HK	991			
Qy	332	KQCKQ	AHLSEDC	DPD---VP	ALHTL	DE-----	AIRLV	NVNSQ	YGOIL-----	Q	373				
Db	992	EQOLKQ	QSIEND	LIEKEN	IQLO	ASQNL	NEQ	QQSQNS	NOLSEK	DQOL	NOLIEK	NQFDQ	KEQ	1051	
Qy	374	MTRKH	LED	TAY---L	VEKMG	QFGVW	SELAN	QAPET	EILFN	SIVQ	VPRI	HEG	NI	SKQD	428
Db	1052	LKQSI	END	LIEKEN	IQLO	ASQNL	NEQ	QQSQNS	NOLSEK	DQOL	NOLIEK	NQFDQ	KEQ	1105	
Qy	429	ETMT	DL	LSI	437										
Db	1106	EQLKQ	QSI	1114											

RESULT	12
T34107	
hypothetical protein C18C4.5 - Caenorhabditis elegans	
C:Species: Caenorhabditis elegans	
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999	
C:Accession: T34107	
R:Gattung, S.	
submitted to the EMBL Data Library, April 1996	
A:Description: The sequence of C. elegans cosmid C18C4.	
A:Reference number: Z21478	
A:Accession: T34107	
A:Status: preliminary; translated from GB/EMBL/DDBJ	
A:Molecule type: DNA	
A:Residues: 1-1091 <GAT>	
A:Cross-references: EMBL:U55369; PIDN:AAC25825.1; GSPDB:GN00023; CESP:C18C4.5	
A:Experimental source: strain Bristol N2; clone C18C4	
C:Genetics:	
A:Gene: C18C4.5	
A:Map position: 5	
A:Introns: 38/3; 95/3; 179/2; 259/3; 301/3; 419/2; 573/3; 613/1; 875/3; 920/2; 9	

Query Match	6.0%;	Score	151.5;	DB	2;	Length	1091;		
Best Local Similarity	20.5%;	Pred. No.	0.095;						
Matches	105;	Conservative	99;	Mismatches	184;	Indels	125;	Gaps	25;

Qy	36	KDKTAISENLKSFSEVG	Idadev	KKALTG	IKOMKIMMER-----	76							
Db	283	KDNNAIL	NVQLREK	GKIDRIQ	VDLLAA	ESRAQ	AEDV	RD	MKRIIT	SKDD	SNLL	341	
Qy	86	-STLUK	KREEKOE	AKLLN	VEOHL	EEER	LCRES	LADSWG	CRSCL	ENNC	MR-----	IY	139
Db	342	QDELR	TEKYQQA	KKNIEND	ETIKOO	ETQIRD	-LGRS	LDEAK	RQ	LRMS	EQR	NEVA	400

Db	1118	KVKECLNIIKSLKELNKEKCONLSDASKUYTELQ-----EIHENLLLKVSDLENYK	1171
Qy	331	EKCQCAHLSEDQDPVAPALHTELDEAIRLVNVSQ-----QYQIOLMTRKHLEPTA	383
Db	1172	KKYEGLDL-EGLKVDVTFQELSKKRDLTFNHESLLRQSAKYKEKLSLASSENKDL	1230
Qy	384	YLVEKMRGQGWSELANQAPETIINSIOVVPRIHE-GNISQDETMMTDLSILPSSN	442
Db	1231	NKVSLSLKQVNELSPKASKVPELE-----RKITNLMHEYSQLGKTTFEDEKRAKALIASRDN	1285
Qy	443	FTLKIPLEESAESSNFIQYVVAKALQHFK	471
Db	1286	BELR-SUKSELSRKRKLEVEYQKVLVEVK	1313
RESULT 14			
B70356			
Chromosome assembly protein homolog - Aquifex aeolicus			
C:Species: Aquifex aeolicus			
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Jun-2000			
C:Accession: B70356			
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O. V.			
Nature 392, 353-358, 1998			
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.			
A:Reference number: A70300; MUID:98196666			
A:Accession: B70356			
A:Status: preliminary; nucleic acid sequence not shown; translation not shown			
A:Molecule type: DNA			
A:Residues: 1-1156 <AQF>			
A:Cross-references: GB:AE000699; NID:g2983238; PIDN:AA06839.1; PID:g2983243; GB:AE000699			
A:Experimental source: strain V5			
C:Genetics:			
A:Gene: xpc			
C:Superfamily: chromosome segregation protein SMC1			
Query Match 5.88; Score 146; DB 2; Length 1156;			
Best Local Similarity 20.08; Pred. No. 0.23;			
Matches 89; Conservative 89; Mismatches 162; Indels 104; Gaps 17;			
Qy	45	LKSPSEVEID-----ADEEVKKALTKQMKIMME-----RKEKEHTNLMSTLKK	90
Db	163	IEETSGEYERKKEKALEAEVLKKEIDTLLIEESNQLKRLKEKEKLEKFKELQ	222
Qy	91	CREKQBALKLLNEVQEHLEEEERLCRE--SLADSWGECRSLENNCRMYTTCOPSWSS	148
Db	223	IKRE--TEAKILLKEKELKKERITNLSLRESLEDITTFIOENKEL-----	271
Qy	149	VKNKIERFFRIYQFLPPHEDNEK---DLPISEKLI--EEDAQLTQMEDVFSQLTVDVNS	204
Db	272	--NERERLLKVNKEMIPFKVKQKFTAEIENAEARSKEKREKELSENVRKNLSELINN	329
Qy	205	LFNRSFNVRQ---MQQEDFTQFSHTSDTDLTPY--FFPAFSKEPMTKADLEOCWDIP	260
Db	330	LLSDKENLEREVGTLQLELEK-----LKREYKSLKEVEREKLRELEEE-----	373
Qy	261	NFFOLFCNF-SVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLITSKMLPGODRGLCGEL	319
Db	374	--ERLKTFIDVKKLEEKLEKTLKNSL-NKEQLELTQIRANLKNKI-----ERI	421
Qy	320	DONLSRCFKPHEKCKQCAHLSEDCPDVPALHTELDEAIRLVNVSQOYQGIOLMTRKHL	379
Db	422	KEDINKLISEREKKEIKKEQEIKRUKAKKKKEEELRNLTQELNIYEKRLSEVRKKL	481
Qy	380	E-----DTAYLVKMRGQFCQWSELAN-QAPE	405
Db	482	EEVLKKGAIEREVRSFSDVDFKDIKGVYGSVELIRVNPEHITALEVAGGRLAKFI	541
Qy	406	----TEITFNISQIOWPRIHEGNIS	425
Db	542	VVEDEEYAKECIQIAKRMNLGRFS	565

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:50:18 ; Search time 17.51 Seconds
(without alignments)
1054.782 Million cell updates/sec

Title: US-09-722-544A-4

Perfect score: 2529

Sequence: 1 MRTWDYNSGNMKPPLLVFI.....FIGYVAKALQHFKEFKTW 477

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	DB ID	Description
1	434	17.2	449	1	CLUS_HUMAN
2	416	16.4	451	1	CLUS_COTJA
3	413.5	16.4	447	1	CLUS_RABIT
4	410.5	16.2	445	1	CLUS_CANFA
5	403	15.9	448	1	CLUS_MOUSE
6	400.5	15.8	439	1	CLUS_BOVIN
7	400.5	15.8	446	1	CLUS_PIG
8	380	15.0	447	1	CLUS_RAT
9	168	6.6	191	1	CLUS_MESAU
10	151	6.0	1526	1	MY52_SCHPO
11	145.5	5.8	594	1	C1K1_YEAST
12	144.5	5.7	1938	1	MYHD_HUMAN
13	142.5	5.6	2663	1	CENE_HUMAN
14	141.5	5.6	3433	1	UTRO_HUMAN
15	141	5.6	1005	1	RA50_METJA
16	137	5.4	1940	1	MYH3_HUMAN
17	136	5.4	2230	1	G0G4_HUMAN
18	135	5.3	1060	1	E651_XENLA
19	134.5	5.3	1957	1	YD86_SCHPO
20	134	5.3	5430	1	ACF7_HUMAN
21	132.5	5.2	852	1	RA50_THEMA
22	131.5	5.2	1679	1	Y109_YEAST
23	130.5	5.2	1940	1	MYH3_RAT
24	129.5	5.1	1427	1	REST_HUMAN
25	129.5	5.1	2077	1	TEGU_HSV6U
26	129	5.1	539	1	MY53_HYDAT
27	129	5.1	1853	1	MY5A_MOUSE
28	129	5.1	1966	1	MY5B_CAEEL
29	129	5.1	2704	1	BPAL_HUMAN
30	128	5.1	1189	1	SCII_CHICK
31	128	5.1	1939	1	MYH1_HUMAN
32	128	5.1	1969	1	MY5A_CAEEL
33	127.5	5.0	1978	1	MYHB_CHICK

ALIGNMENTS

```
RESULT 1
CLUS_HUMAN
ID CLUS_HUMAN STANDARD; PRT; 449 AA.
AC P10309; P11380; P11381;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Clusterin precursor (Complement-associated protein SP-40,40)
DE (Complement cytotoxicity inhibitor) (CLI) (NAL and NA2) (Apolipoprotein
DE J) (Apo-J) (TRPM-2).
GN CLU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=89251601; PubMed=2721499;
RA Kirszbaum L., Sharpe J.A., Murphy B., D'Apice J.F.A., Classon B.,
RA Hudson P., Walker I.D.;
RT "Molecular cloning and characterization of the novel, human
RT complement-associated protein, SP-40,40: a link between the
RT complement and reproductive systems.";
EMBO J. 8:711-718(1989).
[2]
RP SEQUENCE OF 2-449 FROM N.A.
RX MEDLINE=89386692; PubMed=2780565;
RA Jenne D.E., Tschopp J.;
RT "Molecular structure and functional characterization of a human
RT complement cytotoxicity inhibitor found in blood and seminal plasma:
RT identity to sulfated glycoprotein 2, a constituent of rat testis
RT fluid.";
Proc. Natl. Acad. Sci. U.S.A. 86:7123-7127(1989).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90344779; PubMed=1974459;
RA de Silva H.V., Harmony J.A.K., Stuart W.D., Gil C.M., Robbins J.;
RT "Apolipoprotein J: structure and tissue distribution.";
Biochemistry 29:5380-5389(1990).
[4]
RP SEQUENCE FROM N.A.
RX MEDLINE=94237156; PubMed=8181474;
RA Wong P., Taillefer D., Lakins J., Pineault J., Chader G.,
RA Tenniswood M.;
RT "Molecular characterization of human TRPM-2/clusterin, a gene
RT associated with sperm maturation, apoptosis and neurodegeneration.";
Eur. J. Biochem. 221:917-925(1994).
[5]
RP SEQUENCE FROM N.A.
RX MEDLINE=93186813; PubMed=7680346;
RA Pineault J.M., Tenniswood M.;
RT "Genomic organization and expression of the rat TRPM-2 (clusterin)
RT gene, a gene implicated in apoptosis.";
J. Biol. Chem. 268:5021-5031(1993).
[6]
```

Q07970 arabidopsis
Q02440 gallus gall
P09KX2 homo sapien
P35579 homo sapien
P02549 homo sapien
P49454 homo sapien
O67124 aquifex aeo
P26386 saccharomyc
P38871 saccharomyc
O51246 borrelia bu
P29616 gallus gall
Q9Y4I1 homo sapien

34 127 5.0 793 1 KATA_ARATH
35 127 5.0 1829 1 MY5A_CHICK
36 126 5.0 1941 1 MYH2_HUMAN
37 126 5.0 1960 1 MYH9_HUMAN
38 126 5.0 2418 1 SPCA_HUMAN
39 125.5 5.0 3210 1 CENF_HUMAN
40 125 4.9 978 1 RA50_AQUAE
41 125 4.9 1790 1 US01_YEAST
42 124.5 4.9 571 1 YHY4_YEAST
43 124.5 4.9 971 1 Y228_BORBU
44 124.5 4.9 1102 1 MYSC_CHICK
45 124.5 4.9 1855 1 MY5A_HUMAN


```

RP SEQUENCE FROM N.A.
RA Blechschmidt K., Rosenthal A.;
RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE OF 61-449 FROM N.A.
RX MEDLINE-92020896; PubMed=1924317;
RA Danik M., Chabot J.G., Mercier C., Benabid A.L., Chauvin C.,
RA Quirion R., Suh M.;
RT "Human gliomas and epileptic foci express high levels of a mRNA
RT related to rat testicular sulfated glycoprotein 2, a purported marker
RT of cell death.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:8577-8581(1991).
RN
RN SEQUENCE OF 140-449 FROM N.A.
RC TISSUE=Liver;
RA Glew M.D., Kirsbaum L., Bozas S.E., Walker I.D.;
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
RN
RN SEQUENCE OF 23-33 AND 228-240.
RX MEDLINE-91265608; PubMed=3154963;
RA Hochstrasser A.-C., James R.W., Martin B.M., Harrington M.,
RA Hochstrasser D.F., Pometta D., Merrill C.R.;
RT "HDL particle associated proteins in plasma and cerebrospinal fluid:
RT identification and partial sequencing.";
RL Appl. Theor. Electrophor. 1:73-76(1988).
RN
RN SEQUENCE OF 23-33; 229-242; 303-317 AND 397-403.
RX MEDLINE-91230083; PubMed=1903064;
RA James R.W., Hochstrasser A.-C., Borghini I., Martin B.M., Pometta D.,
RA Hochstrasser D.F.;
RT "Characterization of a human high density lipoprotein-associated
RT protein, NAL/NA2. Identity with Sp-40,40, an inhibitor of complement-
RT mediated cytotoxicity.";
RL Arterioscler. Thromb. 11:645-652(1991).
RN
RN SEQUENCE OF 23-52 AND 228-257.
RX MEDLINE-90354412; PubMed=2387851;
RA de Silva H., Stuart W.D., Park Y.B., Mao S.J.T., Gil C.M.,
RA Wetterau J.R., Busch S.J., Harmony J.A.K.;
RT "Purification and characterization of apolipoprotein J.";
RL J. Biol. Chem. 265:14292-14297(1990).
RN
RN SEQUENCE OF 23-37 AND 228-242.
RX MEDLINE-90097955; PubMed=2601725;
RA Choi N.H., Mazda T., Tomita M.;
RT "A serum protein Sp40,40 modulates the formation of membrane attack
RT complex of complement on erythrocytes.";
RL Mol. Immunol. 26:835-840(1989).
RN
RN SEQUENCE OF 23-41 AND 228-246.
RX MEDLINE-93319521; PubMed=8328966;
RA Ghiso J., Matsubara E., Koudinov A., Choi-Miura N.-H., Tomita M.,
RA Wisniewski T., Frangione B.;
RT "The cerebrospinal-fluid soluble form of Alzheimer's amyloid beta is
RT complexed to Sp-40,40 (apolipoprotein J), an inhibitor of the
RT complement membrane-attack complex.";
RL Biochem. J. 293:27-30(1993).
RN
RN PARTIAL SEQUENCE, DISULFIDE BONDS, AND CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE-92201397; PubMed=1551440;
RA Kirsbaum L., Bozas S.E., Walker I.D.;
RT "Sp-40,40, a protein involved in the control of the complement
RT pathway, possesses a unique array of disulphide bridges.";
RL FEBS Lett. 297:70-76(1992).
RN
RN DISULFIDE BONDS.
RX MEDLINE-93147076; PubMed=1491011;
RA Choi-Miura N.H., Takahashi Y., Nakano Y., Tobe T., Tomita M.;
RT "Identification of the disulfide bonds in human plasma protein
RT Sp-40,40 (apolipoprotein-J).";
RL J. Biochem. 112:557-561(1992).
RN
RN CARBOHYDRATE-LINKAGE SITES.
RX

```

```

RX MEDLINE-97477902; PubMed=9336835;
RA Kapron J.T., Hilliard G.M., Lakins J.N., Tenniswood M.P., West K.A.,
RA Carr S.A., Crabb J.W.;
RT "Identification and characterization of glycosylation sites in human
RT serum clusterin.";
RL Protein Sci. 6:2120-2133(1997).
CC
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (APOPTOSIS).
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X14723; CAA32847.1; -
CC EMBL; M25915; AAA35692.1; ALT_INIT.
CC EMBL; J02908; AAAS1765.1; ALT_INIT.
CC EMBL; M74816; AAA60321.1; -
CC EMBL; M64722; AAB06508.1; -
CC EMBL; L00974; AAA60567.1; -
CC EMBL; AF311103; -; NOT_ANNOTATED_CDS.
CC EMBL; M63379; AAB06507.1; -
CC EMBL; M63376; AAB06507.1; JOINED.
CC EMBL; M63377; AAB06507.1; JOINED.
CC EMBL; M63378; AAB06507.1; JOINED.
CC EMBL; A21577; CAA01560.1; -
CC PIR; A35833; A35833.
CC PIR; A37816; A37816.
CC PIR; B37816; B37816.
CC PIR; A41386; A41386.
CC PIR; PLO135; PLO135.
CC PIR; PLO136; PLO136.
CC PIR; S04662; S04662.
CC PIR; S07433; S07433.
CC PIR; S09339; S09339.
CC PIR; S34056; S34056.
CC PIR; S43646; S43646.
CC SWISS-2DPAGE; P10909; HUMAN.
CC InterPro: IPR000753; Clusterin.
CC Pfam: PF01093; Clusterin; 1.
CC SMART; SM00035; CLa; 1.
CC SMART; SM00030; CLb; 1.
CC PROSITE; PS00492; CLUSTERIN_1; 1.
CC PROSITE; PS00493; CLUSTERIN_2; 1.
CC Apoptosis; Complement pathway; Glycoprotein; Plasma; Sulfation;
KW Signal.
FT SIGNAL. 1 22
FT CHAIN 23 449 CLUSTERIN.
FT CHAIN 23 227 BETA-CHAIN.
FT CHAIN 228 449 ALPHA-CHAIN.
FT CHAIN 102 313 INTERCHAIN.
FT DISULFID 113 305 INTERCHAIN.
FT DISULFID 116 302 INTERCHAIN.
FT DISULFID 121 295 INTERCHAIN.
FT DISULFID 129 285 INTERCHAIN.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .).
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .).
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .).
FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .).
FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .).
FT CONFLICT 28 28 D -> S (IN REF. 9 AND 10).
FT CONFLICT 47 47 G -> H (IN REF. 11).
FT CONFLICT 52 52 G -> Q (IN REF. 11).
FT CONFLICT 305 305 C -> M (IN REF. 10).

```



```
Query Match 17.2%; Score 434; DB 1; Length 449;
Best Local Similarity 25.8%; Pred. No. 1.8e-19;
Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

QY 12 MKPPLVIFVCLLWLDKSHCAPTWK-----DKTAISENLKSFSEGEIDAEVKKALT 65
DB 1 MKMTLLFLVGLLL-----TWESQVLGDQTVSDNQLQMSNGSKYVKNKIQAVN 51
QY 66 GIKOMIKMERKEHTNLMTLKKCKEAKLLNEVQHEERLCRESLADSWG 125
DB 52 GVKOIKTLIEKTEERTLLSNLEAKKEDALNETRESETKLKGELPGVCNMTMMALWE 111
QY 126 ECRSLENNCMRIYT-TCQPSWSWKNKIERFFKRIYQFLPFPHEDNEKDLPISEKILIE 184
DB 112 ECKPCLKOTCMKFYARVCRSGSLVGRQLEEFNLQSSPFYFMNGDR-----IDSLLEN 165
QY 185 DAOLQTMEDV----FSQLTVDVNSLNRNFRVQMOQEFQDFQSFHIS-DTDLTPEYF 239
DB 166 DROQTHMLDVMDHFSRASIIIDLFODRF-----FTREPQDTY--HYLPFSLPHRRPHF 218
QY 240 FPAFSKEPTMKADLE-OCWDIPNFQFCNFVSIVSVSETITKMLKATEDLPKQDKAP 298
DB 219 F--PPKSRIVRSLMPPSPYEPLEPHAMFPFLEMIHEA-----QQAMDIHFHSPAF 267
QY 299 DHGLISKMLPGQDRLGCLGDLQNLSCRFKHEKCKQCAHLSDC-----PDVPAHTEL 354
DB 268 QHPPTPIREGDDRTVCREIRHNSTGCLRMKQCDCKREILSVDCSTNNPSQAKLRREL 327
QY 355 DEARLVNVSNOQYQGLQWTRKHELTAYLVKMRGFCGWSELANOAPETELIFNSIQ 414
DB 328 DESQVAERUTRYKNEULLSYQWMLNLTSSLEQLNQFNQVNSRLANLTQGEQYLYLRT 387
QY 415 VVPRHBNISKODETMMTDLSTLPSSNFTLKIPLEBSAESSNFIVGVVAKALQHF-KEH 473
DB 388 TVAS-HTSDSDVSGVTEVVVKLFDSQPIVTVVPEVSRKNPKFMETVAEKALQETRKKH 446

RESULT 2
ID CLUS_COTJA STANDARD; PRT; 451 AA.
AC P14018;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Clusterin precursor (51.5 kDa protein).
GN T64.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroretina;
RX MEDLINE=89239492; PubMed=25411393;
RA Michel D., Gillet G., Volovitch M., Pessac B., Calothy G., Brun G.;
RT "Expression of a novel gene encoding a 51.5 kD precursor protein is
RT induced by different retroviral oncogenes in quail neuroretinal
RT cells."
RL Oncogene Res. 4:127-136(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95262670; PubMed=7744033;
RA Michel D., Chatelain G., Herault Y., Brun G.;
RT "The expression of the avian clusterin gene can be driven by two
RT alternative promoters with distinct regulatory elements."
RL Eur. J. Biochem. 229:215-223(1995).
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH.
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER
```

QY 435 LSILPSSNFTLKIPLESSESSNFGYVAKALQHFKEH 473
 DB 409 VOLFDSEPLSLTPGDISWDDPRFMEIYVAEQALQHYKON 447

RESULT 3
 CLUS_RABIT STANDARD; PRT; 447 AA.

AC Q9XSC5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Clusterin precursor (Apolipoprotein J) (Apo-J).
 GN CLU.

OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAPANESE WHITE;
 RA Miyata M., Biro S., Kaleda H., Eto H., Orihara K., Kihara T.,
 RA Obata H., Matsushita N., Matsuyama T., Tei C.;
 RT "Apolipoprotein J/clusterin is induced in vascular smooth muscle cells
 by balloon injury.";
 RL Submitted (JAN1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
 CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
 CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
 CC CELL DEATH (BY SIMILARITY).
 CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
 CC -!- SUBCELLULAR LOCATION: PRESENT IN CHROMAFFIN GRANULES.
 CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL: AF118852; AAD24461.1; -
 DR InterPro; IPR000753; Clusterin.
 DR Pfam; PF01093; Clusterin; 1.
 DR SMART; SM00035; CLA; 1.
 DR SMART; SM00030; CLB; 1.
 DR PROSITE; PS00492; CLUSTERIN_1; 1.
 KW Glycoprotein; Plasma; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 447
 FT CHAIN 23 447
 FT CHAIN 23 447
 FT CHAIN 226 447
 FT DISULFID 101 311
 FT DISULFID 112 303
 FT DISULFID 115 300
 FT DISULFID 120 293
 FT DISULFID 128 283
 FT CARBOHYD 85 85
 FT CARBOHYD 102 102
 FT CARBOHYD 144 144
 FT CARBOHYD 289 289
 FT CARBOHYD 326 326
 FT CARBOHYD 352 352
 FT CARBOHYD 372 372
 SQ SEQUENCE 447 AA; 51850 MW; COCA338EEB08590A CRC64;

Query Match 16.4%; Score 413.5; DB 1; Length 447;
 Best Local Similarity 24.1%; Pred. No. 3e-18;
 Matches 111; Conservative 93; Mismatches 204; Indels 53; Gaps 11;

QY 34 TWKDKTAISENLKSFSEVGIDA-----DEEVKALTIQKIMMERKEHTNLMST 87

Db 13 SWRGQVLGDQLVSDNELQEMSTQSGKYDREIONAVKGVEIKTLIEKTEERKTLISV 72
 QY 88 LKCKREEKQEAALKLINEVQEHLEBEERLCRESLADSWGECRSCLENNCMRIYT-TCOPSW 146
 Db 73 LEEAKKNKEDALNETRDSSETKLKAPPEVCNETMMAWEECKPCLKQTCMKFYARVCRSGS 132
 QY 147 SSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLIBEDAQLTQMEDVFSOLTVDVNSLF 206
 Db 133 GLVGRQLREELNQSSPFYFWINGDR-----IDSLLENDROQSHVLDV-----MODSF 179
 QY 207 NRSFNVFQMOQERDQTFQSHFISDTLTETEPYFFPAESKEPMTKADLQCWDIP----- 260
 Db 180 NRATGI---MDLFDQDFRTHKPDQTFYHSPFSY--FRPPPLHYAKSLVRNIMPLSLYG 234
 QY 261 --NFFQLCFNFSVSIYESVSETITKMLKAIEDLPKQD---KAPDHGGLISKMLPGQDRL 315
 Db 235 PLNQDFEQPFEMTHQAQQAQDVHLHSPAYQTPNVEITGGPD-----DRAV 282
 QY 316 CGELDONLSRCFKFHEKCKQCOAHLSEDC----PDVPALHTELDRAIRLVVNSNQYQGI 371
 Db 283 CKETRHNSGTCLRMKQCAKQCEILSDVCSANNPSQNLQELNDSLRLABELTKRYNEL 342
 QY 372 LQMRKHLIEDTAYLVEKMRGQGVSELANQAPETETIIFNSIQVVPRIHGNISKQDET 431
 Db 343 LQSYQMKMLNTSSLLDQPNQFNWVSQANLTQGPQDYILRVSTVTS--HTSESEAPSRV 401
 QY 432 MTDLSILPSSNFTLKIPLESSESSNFGYVAKALQHFKE 472
 Db 402 EVVYKLFDSDDPTITIPPEVSRDNPKNFETVAERKALQEVYK 442

RESULT 4
 CLUS_CANFA STANDARD; PRT; 445 AA.

AC P25473;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Clusterin precursor (Glycoprotein 80) (GP80).
 GN CLU.

OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91236776; PubMed=2033078;
 RA Hartmann K., Rauch J., Urban J., Parczyk K., Diel P., Pillarsky C.,
 RA Appel D., Haase W., Mann K., Wellner A., Koch-Brandt C.;
 RT "Molecular cloning of gp 80, a glycoprotein complex secreted by
 RT kidney cells in vitro and in vivo. A link to the reproductive system
 RL and to the complement cascade.";
 RL J. Biol. Chem. 266:9924-9931(1991).
 CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
 CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
 CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
 CC CELL DEATH.
 CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
 CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL: M55251; AAA30846.1; -
 DR InterPro; IPR000753; Clusterin.
 DR Pfam; PF01093; Clusterin; 1.

```
DR SMART: SM00035; Cls: 1.
DR SMART: SM00030; Clb: 1.
DR PROSITE: PS00492; CLUSTERIN_1; 1.
DR PROSITE: PS00493; CLUSTERIN_2; 1.
DR Glycoprotein: Plasma; Signal.
FT SIGNAL 1 22
FT CHAIN 23 445
FT CHAIN 23 445
FT CHAIN 23 445
FT CHAIN 23 445
FT DISULFID 102 309
FT DISULFID 113 301
FT DISULFID 116 298
FT DISULFID 121 291
FT DISULFID 129 281
FT CARBOHYD 86 86
FT CARBOHYD 103 103
FT CARBOHYD 145 145
FT CARBOHYD 277 277
FT CARBOHYD 287 287
FT CARBOHYD 350 350
FT CARBOHYD 370 370
SQ SEQUENCE 445 AA; 51789 MW; 023A37266ABEF374 CRC64;

Query Match 16.2%; Score 410.5; DB 1; Length 445;
Best Local Similarity 23.9%; Pred. No. 4.6e-18;
Matches 117; Conservative 102; Mismatches 192; Indels 79; Gaps 15;

QY 12 MKPPLLVFIVCLWLKSHCAPTKDKTAISENLKSFSEVGEIDAD-----EVVKALT 65
DB 1 MKMTLLLVGLL-----TWNGRVLGDAQVSDTEIQEMSTEGSKYINKEINALK 51
QY 66 GIKOMTKMERKKEHTNMLTKKREEKQOEALKILNEVQEHLEERLCRESLADSWG 125
DB 52 GVQIKTLIEQTNEERKSLNLEEAKKKEDALNTKQSETKLKASQGVNCFMFWALWE 111
QY 126 ECRSCLENNCRMYT-TCQPSWSVNKNIRFFRKIYQFLPPHEDNEKDLPISEKLIEE 184
DB 112 ECKPCLKQTCMKFYARVCRSGSLGVHGLEEFNLQSSPFYFWNGDR-----IDSILEN 165
QY 185 DAQLTQMEDVFSOLTVDVNSLFNRNFVFMQOEQDQFQSHFIS-----DTDLTPEYF 240
DB 166 DROOTHALDV-----MQDSFRASSI-----MDLFDQFFTFREPQDTHYSP--F 209
QY 241 PAFSKEPMTKADLEQCWDI-----PNFFQLCFNFSVSIYESVETITKMLKAIE--- 289
DB 210 SLFQRRPFPNPKFRIARNIIPFPFQPLNPHDMFQPFEDMIHQ-A-QQAMDVNLHRIPYHF 268
QY 290 --DLPKQDKAPDHGGLISKMLPQDRLGCLGELDONLSRCFKFHEKCKQQAHLSEDC--- 344
DB 269 PIEFPEED-----NRTVCKEIRHNSTGCLMKMDQCEKQCEILSVDCSSN 312
QY 345 -PDVPALHTLDEAIRLVNVSNOQYGOILQMTKRKLEDYALVEKMRGQFGWSELANOQA 403
DB 313 NPAQVQLRQELSNLQIAEKFTKLYDELQSYQEKMFNTSSLKQLQNEQFSWVSQLANLT 372
QY 404 PETEITFNISQVY-PRHEGNTISKQDETMMTDLISLPSSNFTLKIPLESAESSNFIGYV 462
DB 373 QSEDPFYLTQVTGTSQSDSNVPVGTQKV--VKLFDSDPITVWIPEAVSRNNPKMETV 430
QY 463 VAKALQHFKE 472
DB 431 AERALQEIYRQ 440

RESULT 5
ID CLUS_MOUSE STANDARD; PRT; 448 AA.
AC Q06890;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Clusterin precursor (Sulfated glycoprotein 2) (SGP-2) (Clusterin)
```

```
DE (Apolipoprotein J) (Apo-J).
GN CLJ OR MSGP-2 OR APOJ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RN SEQUENCE FROM N.A.
RP TISSUE=Testis;
RC MEDLINE=93356785; PubMed=8352774;
RX Lee K.-H., Ji Y.-M., Lim H.-M., Lee S.-C., You K.-H.;
RA "Molecular cloning and sequencing of sulfated glycoprotein-2 cDNA
RT from testis of mouse: implications of two different mRNAs of SGP-2.";
RL Biochem. Biophys. Res. Commun. 194:1175-1180(1993).
RN 12
RN SEQUENCE FROM N.A.
RP MEDLINE=93359508; PubMed=8354695;
RX French L.E., Chonn A., Ducrest D., Baumann B., Belin D., Wohlwend A.,
RA Kiss J.Z., Sappino A.P., Tschopp J., Schifferli J.A.;
RT "Murine clusterin: molecular cloning and mRNA localization of a gene
RT associated with epithelial differentiation processes during
RT embryogenesis.";
RL J. Cell Biol. 122:1119-1130(1993).
RN 13
RN SEQUENCE FROM N.A.
RP Hodgdon B.A., Min B.H., Yan H., Farris J.A., Foster D.N.,
RA Strauch A.R.;
RL Submitted (XX-1992) to the EMBL/GenBank/DBJ databases.
RN 14
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RX MEDLINE=94223204; PubMed=8169523;
RA Jordan-Starck T.C., Lund S.D., Witte D.P., Aronow B.J., Ley C.A.,
RA Stuart W.D., Swerffeger D.K., Clayton L.R., Sells S.F., Paigen B.;
RT "Mouse apolipoprotein J: characterization of a gene implicated in
RT atherosclerosis.";
RL J. Lipid Res. 35:194-210(1994).
CC -!- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A
CC VARIETY OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS,
CC MEMBRANES, AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED
CC WITH PROGRAMMED CELL DEATH.
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC -!- TISSUE SPECIFICITY: MOST ABUNDANT IN STOMACH, LIVER, BRAIN, AND
CC TESTIS, WITH INTERMEDIATE LEVELS IN HEART, OVARY, AND KIDNEY.
CC -!- PTM: INTENSIVELY GLYCOSYLATED WITH SULFATED N-LINKED
CC CARBOHYDRATES.
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D14077; BAA03162.1; -
CC EMBL: L08235; AAA37422.1; -
CC EMBL: L05670; AAA37284.1; -
CC EMBL: S70244; AAB30623.1; -
CC PIR: A40714; A40714.
CC PIR: JN0699; JN0699.
CC MGI: 88423; Clu.
CC InterPro: IPR000753; Clusterin.
CC Pfam: PF01093; Clusterin; 1.
CC SMART: SM00035; Cls; 1.
CC SMART: SM00030; Clb; 1.
CC PROSITE: PS00492; CLUSTERIN_1; 1.
CC PROSITE: PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Sulfation; Signal.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 448
FT CLUSTERIN.
FT BETA-CHAIN (SMALL SUBUNIT)
FT CHAIN 22 226
```



```
Db 233 -----NFHDVTFQPF-----YDHIHQQAQMDAHLQ-----RTPYHFPTM-EFTENNDRTVC 277
QY 317 GELDONLSRCFKHEKQKQAHLSDEC-----PDVPAHTELDIAIRLVNVNQYQOIL 372
Db 278 KEIRHNSGTGLRMKQDQKCEQLEIVDCSASNTQTLLRQOLNASLQAEKFSRLYDQLL 337
QY 373 QMTRKHLEDIAYLVKMRGQGWVSELANOAPETE-----IIFNSQVVPRIHEGNISK 426
Db 338 QSYOQKMLNTSALLKQLENEQFTWVSQLANLTQSDQHYLVFTVNSHNSDPSIPSG---- 393
QY 427 QDETMTDLSLTPSSNFTLPILEESAESSNFCYVYVAKALQHF 471
Db 394 -----LTKVIVKLFNSPPIITVTPQEVSSPFMENVAEKALQOYR 433

RESULT 7
CLUS_PIG STANDARD; PRT; 446 AA.
AC Q29549;
DT 13-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Clusterin precursor (Complement cytotoxic inhibitor) (CLI).
GN CLI.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92184774; PubMed=1544909;
RA Diemer V., Hoyte M., Baglioni C., Millis A.J.;
RT "Expression of porcine complement cytotoxic inhibitor mRNA in
RT cultured aortic smooth muscle cells. Changes during differentiation
RT in vitro."
RL J. Biol. Chem. 267:5257-5264(1992).
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN BRAIN AND LIVER; LOWER
CC LEVELS ARE DETECTED IN OTHER TISSUES, INCLUDING THE AORTA.
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M84639; AAA31013.1;
CC InterPro: IPR000753; Clusterin.
CC Pfam: PF01093; Clusterin; 1.
CC SMART: SM00035; CLA; 1.
CC SMART: SM00030; CLB; 1.
CC PROSITE: PS00492; CLUSTERIN_1; 1.
CC PROSITE: PS00493; CLUSTERIN_2; 1.
CC Glycoprotein; Signal
KW SIGNAL 1 28 BY SIMILARITY.
FT CHAIN 29 446 CLUSTERIN.
FT CHAIN 29 446 BETA-CHAIN (A CHAIN).
FT CHAIN 228 446 ALPHA-CHAIN (B CHAIN).
FT FT DISULFID 102 312 INTERCHAIN (BY SIMILARITY).
FT FT DISULFID 113 304 INTERCHAIN (BY SIMILARITY).
FT FT DISULFID 116 301 INTERCHAIN (BY SIMILARITY).
FT FT DISULFID 121 294 INTERCHAIN (BY SIMILARITY).
FT FT DISULFID 129 284 INTERCHAIN (BY SIMILARITY).
FT FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 446 AA; 51774 MW; 51774 MW; BLD5B434B668E3AA CRC64;

Query Match 15.88; Score 400.5; DB 1; Length 446;
Best Local Similarity 24.4%; Pred. No. 1.9e-17;
Matches 118; Conservative 98; Mismatches 196; Indels 71; Gaps 16;

QY 16 LLVPIVCLMLKDSHCAPTWK-----DKTAISENLKSFSEVEIGEDADAEVKALTGK 68
Db 4 LLLLVGLLL-----TWENGPWVLGDKAISDKLOEMSTEGSKYVNEIKNALKEV 54
QY 69 QMKIMMERKEKHTNLMSTLKKREEKQEAALKLLNEVOEHLERLEERLCRLSADSGECR 128
Db 55 QIKTLIEQSNERRKSLSSLEAEAKKEDALNDTRDTETKLGSGQLCNETMMALWEEK 114
QY 129 SCLENNCMRIYT-TCOPSSWSSVKNKTERFFRKIVQFLPFPHEDNEKDLPISEKLIEDAQ 187
Db 115 PCLQKTCMKFYARVCRSGSLVGHQLEEFNLQSSPFFYFWINGDR-----IDSLMENDRQ 168
QY 188 LTQMEDVFSQLTVDVANSLENRSFNFRMQOQEPDQTFQSHFIS-----DTDLTEPY----- 238
Db 169 QSHVMDI-----MEDSFNRASNI-----MDLQDFQDFRFPDPTOFFSFGSSHR 214
QY 239 ---FFPAFSKEPMKADLEOCWDIPNPFQFCNFSVSIYESVETITKMLKAIEDLPKQD 295
Db 215 GSLFFNPKSRFARNIMPFLFTDL-NYHDMFQPF---FDMIHQAQQAQMDAHLRIPYH- 268
QY 296 KAPDHGSLISKMLP--GQDRGLCGELDONLSRCFKHEKQKQAHLSDEC-----PDVPA 349
Db 269 -FPEAG-----VPENSNDRAVCKEIRHNSGTCLRWKQDCEKREILSSASNSQMQ 321
QY 350 LHTELDEAIRLVNVNQYQOILQMTKRLHLEDIAYLVKMRGQGWVSELANOAPETEII 409
Db 322 LRQELYSLQMAEKFKSLYDQLLSYQOQKMLNTSSLLKQLNEQFSWVSQLANTQNDRY 381
QY 410 FNSIQVPRHIEGNISKQDTEMTDLSILPSSNFTLPILEESAESSNFTGYVYVAKALQH 469
Db 382 YLVQVTV-NSHGSDPSVPSGLTKRVVVKLEDSYPITLIIPQEV--DPKFMTVAEEALQ 438
QY 470 FKE 472
Db 439 YRQ 441

RESULT 8
CLUS_RAT STANDARD; PRT; 447 AA.
AC P0537L;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Clusterin precursor (Sulfated glycoprotein 2) (SGP-2) (Dimeric acid
DE glycoprotein) (DAG) (Testosterone repressed prostate message-2) (TRPM-
DE 2).
GN CLI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-45 AND 227-241.
RX MEDLINE=88000523; PubMed=3651384;
RA Collard M.W., Griswold M.D.;
RT "Biosynthesis and molecular cloning of sulfated glycoprotein 2
RT secreted by rat Sertoli cells."
RL Biochemistry 26:3297-3303(1987).
RN [2]
RP SEQUENCE FROM N.A.
```



```
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_Tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW Multigene family.
FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
FT DOMAIN 785 814 IQ.
FT DOMAIN 843 1938 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 659 681 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 761 775 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 130 130 METHYLATION (SH-1) (POTENTIAL).
FT MOD_RES 699 699 ALKYLATION (SH-2) (POTENTIAL).
FT MOD_RES 709 709 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1938 AA; 223678 MW; 1F6D006416381CD5 CRC64;

Query Match 5.7%; Score 144.5; DB 1; Length 1938;
Best Local Similarity 20.7%; Pred. No. 0.29;
Matches 100; Conservative 93; Mismatches 206; Indels 83; Gaps 17;

QY 41 ISENLKSFSEVGEIDAEV-KKALTCIKOMKIMMERKEHNLMTLKKCR-----EEK 95
DB 1150 ISELEAS--GATSAIQEMKKREAEFKMRDRLEATLQHEATAATLKKQADSVAEL 1207

QY 96 QEALKLLNEVQEHLEEBERLCRESLAD--SWGECRSLENNCKRIVTTCOPSSVKNKI 153
DB 1208 GEQIDNLRQVKQKLEKSELKMEIDDMASNIALSCKSNIRTCRTVEDQFSEIRAKD 1267

QY 154 ERPRKIYQFLPFHEDNEKDLPISEKLEEDAQLTQMEDVFSQLVVDVNSLFNRSFNVF 213
DB 1268 EQQTQLI-----HDLMQKARLOTQNGELSHRVEEKESLIISQATKSKQALTQLEELK 1320

QY 214 RMOQOEFD-QTFQSHFTS-----DTDLTPEYFFPAFKEPTKADLQCV-----DIPNFF 263
DB 1321 ROMEETKAKNAHAHALQSSRHDCDLR-----EQYEEQEAQALORALSKANSEVAQWK 1376

QY 264 QLCNFSVSIYESVETITKMLKAIEDL-PKQKAPDHGLISK-----MLPGQDRGLGEL 319
DB 1377 TKYETDAIQRTTELEEAKKLAQRLQEAEEKTETANSKCSASLEKTKORLQGEVEDLMDL 1436

QY 320 DONLSRC-----KFEHEKCKQQAHLSEDCPDVPALHTLDEAIRLVNVS 365
DB 1437 ERSHTACATLDKKQRFDKVLAEWKOKLDESQAELSAQKESLSLTEL---FKMRNAYE 1493

QY 366 QYQGI--LQMTRKHLEDYAYLVKMRGFGWYSELANOAPET-----ELI 409
DB 1494 EYVDQLETLRRKNLQEE-----ISDLTEQIAETGKNLQEAETKKLVEQE 1540

QY 410 FNSIQVPRHGENISKQDETWM---TDLISLPSSNFTTKIPLESAESNFIYGVVAKA 466
DB 1541 KSDIQVALBEVGSLEHEESKILRVOLSELQVSKSELDKRVIEKDEIEQLKRNQRAAEA 1600

QY 467 LQ 468
DB 1601 LQ 1602

RESULT 13
CENE_HUMAN
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224.
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
```

```
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and HUBB1.";
RL J. Cell Biol. 143:49-63(1998).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z15005; CAA78727.1;
DR PIR; S28261; S28261.
DR HSP; P17119; 3KAR.
DR MIM; 117143;
DR InterPro; IPR001752; kinesin.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere.
KW DOMAIN 1 335 KINESIN-MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8B8 CRC64;

Query Match 5.6%; Score 142.5; DB 1; Length 2663;
Best Local Similarity 21.5%; Pred. No. 0.56;
Matches 111; Conservative 81; Mismatches 191; Indels 133; Gaps 23;

QY 36 KDKTAISENLKSFSE-----VGEIDA-----DEEVKALTGKQMK----- 71
DB 1200 KKKVKLQKQSFETDRHLRGYREIATGLQTKELKIAHILKHEQETIDELRRSVS 1259
QY 72 -----IMMERKEHNTNLMSTLKKREEKQEAAL---KLLNEVQEHLEEBERLCRESLA- 121
```


Db 1551 KEAASLSENLSTETELVOKSTSEGLLDGDLDTETISWAKNVKLDLE-----KRAADLNTTIT 1605
QY 87 TLKCKREERQKALKLINEVQHELEERLCRESLADSWGECRSCLNNCMRWITTCQPSW 146
Db 1606 -----ESSAALQNLIESEPI-LEERLC--VLNAGWSRVRTWTDWCNTL----- 1647
QY 147 SVVKNKIERFFKI-----YQFLPFPHEDNEKDLPISEKIE-----EDAQLTQMED 193
Db 1648 MHQNOLEIFDGNVAHISTWLYQAEALDEIEKKPTSKOEIVKRLVSELDDANL-QVEN 1706
QY 194 VFSQTLVDVNSLNFNSFVRMQQEFDTQF--SHFISDTDLTETPYFFPAPSKPEMTKA 251
Db 1707 VDDQALILNAGSSRELVPEKLAELNRFKVSQHIKSAKL-----LIAQEP----- 1755
QY 252 DLEQCDWIDPFFQFCNFSVSVSEITITKMLKAI-----EDLPKQDAPDHGGGLISM 307
Db 1756 -LYQCLVTTTETFTGVFP--SDLEKLENDIENMLKEVEXHLESSDEDEKMDSESAQIEV 1812
QY 308 LFGQDRGLGELDONLSRCFKFHEKQKQAHLSDECPDPVPAHTELDRAIRLVNVSNOQ 367
Db 1813 LQGEEMLQHPMEDN-----KKEKIRLQLL-----LLHTRYNK-IRKAIPIQQRK 1855
QY 368 YGQIQLQMKHLEDYAYLVEKMRGQGVSELANQAPETEIFNSIQVVPRIHEGNISQK 427
Db 1856 MGOLASGIRKSLPTDYLVE-----INKILLCDDVVELSLNVPDELNTAIYE-DFSFQ 1906
QY 428 DETM 431
Db 1907 EDSL 1910

RESULT 15
ID RA50.METJA STANDARD; PRT; 1005 AA.
AC Q58718;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR M31322.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii";
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mrell complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mrell by unwinding
CC and/or repositioning DNA ends into the mrell active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mrell (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: U67572; AAB99331.1; -;
DR TIGR: M31322; -;
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR001687; ATP_GTP_A.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02483; SMC_C; 1.
DR Pfam: PF02463; SMC_N; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 32 39 ATP (BY SIMILARITY).
FT DOMAIN 158 849 COILED COIL (POTENTIAL).
SQ SEQUENCE 1005 AA; 119387 MW; 9BBB48173E788F3 CRC64;

Query Match 5.6%; Score 141; DB 1; Length 1005;
Best Local Similarity 19.8%; Pred. No. 0.22;
Matches 90; Conservative 80; Mismatches 172; Indels 112; Gaps 17;

QY 20 IYCLLWLDKSHCAPTKWKDTAISENLKSPSEVGEIDAEVKKALGIKQMKIMMERKEK 79
Db 137 IAKFLSLKPESEKLETYAKVLIGIDEFKCYQKMGEL--VKEYERLERIEGELNYKENYEK 194
QY 80 EHTNLMSTLKKCKREEKQKALKLLNEVQHELEERERESLADSWGECRSCLNNCMRIY 139
Db 195 ELKNMSQLEERKNKLMKLNKLNKIKKFEDEKLFNE----- 233
QY 140 TTCQPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISE---KLIEEDAQLTQMEDVFS 196
Db 234 -----WENKLLYKFKINKL-----EERKRALEKLNQELKILEYDL----- 269
QY 197 QLTVDVNSLNFNSFVRMQQEFQ-----TFQSHFISDTDLTETPYFFPAPSKPEP 247
Db 270 NTVVEARETLNRHKDEYKYSLVDEIRKIESRLKLSHYEDYKLTGK-----QLE 321
QY 248 MTKADLEOCWDIPNFFQFCNFSVSVSEITITKMLKAIED-LPKQDKAPDHGGGLISK 306
Db 322 ILKGDIEKLKEFIN-----KSKYRDDIDNLDLTLNKKIKDEIERVETIKD---LLEE 369
QY 307 MIPGQDRGLGELDONLSRCFKFHEKQKQAHLSDECPDPVPAHTELDRAIRLVNVSNOQ 366
Db 370 L-----KNLNEEIEKIEKYKICECKEYK-----YLELEKAVEYNKLT 412
QY 367 QYQIQLQMTF---KHLEDYAYLVEKMRGQGVSELANQAPETEIFNSIQV---PRIH 420
Db 413 EYITLQEKKSIEKNINDLETINKL-----LEETKN--IDIESIENSLKEIEKKKVL 464
QY 421 EGNISKQDETMTDLSILPSSNFTLKIPLIESAE 454
Db 465 E-NLQEKIELNKKKGEINSEIKRLAKILDELKE 497

Search completed: July 2, 2002, 11:50:20
Job time: 379 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:43:23 ; Search time 56.55 Seconds
(without alignments)
936.909 Million cell updates/sec

Title: US-09-722-544A-4
Perfect score: 2529
Sequence: 1 MRTWDYNSGNMKPPLLVFI.....FIGYVAKALQHEKFKTW 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2489	98.4	477	20 AAY30786	Protein encoded by
2	2464	97.4	466	20 AAY30793	Immature human HKN
3	2441	96.5	495	20 AAY30785	Protein encoded by
4	2350	92.9	446	20 AAY30792	Mature secreted hum
5	1757.5	69.5	465	20 AAY30791	Bovine HKNG1 ortho
6	1639	64.8	466	20 AAY30787	Protein encoded by
7	1558	61.6	450	20 AAY30788	Guinea pig HKNG1 o
8	1277	50.5	374	20 AAY30789	Guinea pig HKNG1 o
9	1272.5	50.3	373	20 AAY30790	Guinea pig HKNG1 o
10	780.5	30.9	521	20 AAY30794	Amino acid sequenc
11	434	17.2	449	22 AAU28048	Novel human secret

12	432	17.1	448	12	AA11704	Cytolysis Inhibito
13	405.5	16.0	416	22	ABB50285	Apolipoprotein J o
14	400.5	15.8	446	21	AA03441	Porcine clusterin
15	320	12.7	363	22	AA03764	Human gene 1 encod
16	173.5	6.9	148	22	AA000607	Human polypeptide
17	160.5	6.3	116	21	AA03745	Human secreted pro
18	159.5	6.3	139	22	AA03783	Human gene 1 encod
19	159.5	6.3	139	22	AA01703	Human gene 4 encod
20	152.5	6.0	247	22	AA03787	Human gene 1 encod
21	145	5.7	1372	19	AAW56473	Protein with Rho p
22	143	5.7	1931	22	ABB61012	Drosophila melanog
23	142.5	5.6	2633	22	ABG06505	Novel human diagno
24	141.5	5.6	3433	18	AAW22017	Utrrophin. Homo sa
25	141	5.6	944	21	AAV67600	Human adipose tiss
26	141	5.6	1388	19	AAW56475	Human adipose tiss
27	139.5	5.5	2663	22	AAW39097	Human polypeptide
28	139.5	5.5	2688	22	AAW40883	Human polypeptide
29	139	5.5	1374	22	AAW69070	Human male enhance
30	136	5.4	934	22	AAU01768	Human secreted pro
31	135.5	5.4	2517	21	AAV71159	Human phosphodiast
32	135.5	5.4	1427	12	AA10534	Human 160kD mediat
33	134	5.3	5373	22	AAU14603	Novel bone marrow
34	134	5.3	5447	22	AAU14697	Novel bone marrow
35	130.5	5.2	2482	16	AAW23996	Human mitosin. Ho
36	130.5	5.2	2482	19	AAW23996	Human mitosin amin
37	130	5.1	990	22	AAW78520	Human protein SEQ
38	130	5.1	1788	22	AAW40467	Human polypeptide
39	128.5	5.1	1392	20	AAW06999	Restin protein seq
40	128.5	5.1	3248	17	AAW99795	Kinetochore protei
41	127.5	5.0	1411	17	AAW02258	Nucleolar/endosoma
42	127.5	5.0	2053	22	AAW03501	Human protein kina
43	127	5.0	1047	18	AAW01535	Cellular homologue
44	126.5	5.0	1780	22	AAW38681	Human polypeptide
45	126	5.0	931	22	AAW79504	Human protein SEQ

ALIGNMENTS

RESULT 1
AAY30786
ID AAY30786 standard; Protein; 477 AA.

AC AAY30786;

DT 23-NOV-1999 (first entry)

DE Protein encoded by a human HKNG1 splice variant HKNG1-V1.

KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW schizophrenia; splice variant.

OS Homo sapiens.

PN WO9947535-A1.

PD 23-SEP-1999.

PF 16-MAR-1999; 99WO-US05606.

PR 16-MAR-1998; 98US-0078044.

PR 05-JUN-1998; 98US-0088312.

PR 28-OCT-1998; 98US-0106056.

PR 22-JAN-1999; 99US-0236134.

PA (MILL-) MILLENNIUM PHARM INC.

PA (REGC) UNIV CALIFORNIA.

PI Chen H, Freimer NB;

XX WPI; 1999-562047/47.

DR N-PSDB; AA210751.

XX New HKNG1 polynucleotides useful in diagnosis and treatment of
PT neuropsychiatric disorders, e.g. bipolar affective disorders and
PT schizophrenia
XX
PS Claim 1; Fig 2A-B; 205pp; English.
XX
CC The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)
CC splice variant. HKNG1 is a gene associated with bipolar affective
CC disorder (BAD). HKNG1 polynucleotides are useful to identify compounds
CC modulating HKNG1 gene expression or HKNG1 polypeptide expression/
CC activity. Compounds inhibiting or enhancing HKNG1 gene expression or
CC activity in individuals can then be administered therapeutically to
CC treat HKNG1-mediated disorders, especially neuropsychiatric disorders
CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as
CC early-onset autosomal dominant myopia. The polynucleotides can be used
CC in gene therapy techniques to treat such disorders. They are also useful
CC in diagnosis to identify individuals having, or at risk of developing,
CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such
CC mutations especially result in the production of a protein with a
CC different sequence to the human full-length HKNG1 polypeptide or
CC splice variant sequences, especially the substitution of a lysine for
CC a glutamic acid at residue 202 or 184. The polynucleotides are also
CC useful in gene mapping, to produce probes or primers to identify
CC similar sequences (e.g. mutants or sequences from different species)
CC and to produce transgenic animals.
XX
SQ Sequence 477 AA;

Query Match 98.4%; Score 2489; DB 20; Length 477;
Best Local Similarity 99.4%; Pred. No. 6.7e-197;
Matches 473; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MRTWDYNSGNKPKPLLYFIVCLLWKDSHCAPTWKDKTAISENLKSFSEGEIDADEEV 60
Db 1 mrtwdysngnmkpllyfivcllwkdshcaptwkdktaisenlksfsegeidadeev 60
Qy 61 KKAALTGIKOMKIMMERKEKEHTNLMSTLKKCKREEKQKALKLNEYQHELEERLCRESL 120
Db 61 kkaaltgikqmkimmerkekehtnlmstlkkckreekqealkllneyqheleerlcresl 120
Qy 121 ADSNGECRSCLENNCMRYTTCQPSWSSVKNKIERFFRKIYQFLPPFHEDNEKOLPISEK 180
Db 121 adswgecrscleenncmrlyttccqpswssvknkierffrkikyqlfpfhednekdlpisek 180
Qy 181 LIEEDAQLTQMEDVFSQLTVDVNSLFRSNNVFRMQQEFQDTFQSHRISDITLPEYFF 240
Db 181 lieedaqltqmedvfsqltvdvnslnfrsnvfrmqgefqtfgskfistdltpeyff 240
Qy 241 PAFSKEPTKADLEQCWDIPNFQLFQNFVSIVESVSETITKMLKATEDLPKQDKAPDH 300
Db 241 pafskeptkadleqcwdipwffqifcnfsvisvesvsetitkmlkaledlpkqdkapdh 300
Qy 301 GGLTSKMLPGODRGICGLDQNLSCFKFHEKCKQKQAHLEDQCDVPALHTLDEARL 360
Db 301 ggllskmlpgqdgrlcgldqnlscfkfmeckqkqahlsedcdvpalhteldearl 360
Qy 361 VNVSNOQYQGIQWTRKHLEDTAYLVEKMRGQFGVWSLANQAPETEIFNSIOWVPRIH 420
Db 361 vnvsngqyqgilqtrkhltdtaylvekmrgqfgvwselanqapeteifnsiqvprih 420
Qy 421 EGNISKQDETMMTDLISLPSSNFTLKIPLESAAESSNIGYVVAKALQHFKEHFT 476
Db 421 egniskqdetmtdlslpssnftlikipleesaessnigyvvakalqhfkehft 476

RESULT 2
AAI30793
ID AAY30793 standard; Protein; 466 AA.
XX
AC AAY30793;
XX

DT 23-NOV-1999 (first entry)
XX
DE Immature human HKNG1 protein form 2.
XX
KW HKNG1: Hong Kong new gene 1; bipolar affective disorder; BAD;
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW schizophrenia; splice variant.
XX
OS Homo sapiens.
XX
PN WO9947535-A1.
XX
PD 23-SEP-1999.
XX
XX 16-MAR-1999; 99WO-US05606.
XX
XX 16-MAR-1998; 98US-0078044.
XX 05-JUN-1998; 98US-0088312.
XX 28-OCT-1998; 98US-0106056.
XX 22-JAN-1999; 99US-0236134.
XX (MILL-) MILLENNIUM PHARM INC.
XX (RESC) UNIV CALIFORNIA.
XX
XX Chen H, Freimer NB;
XX
XX WPI; 1999-562047/47.
XX
XX New HKNG1 polynucleotides useful in diagnosis and treatment of
PT neuropsychiatric disorders, e.g. bipolar affective disorders and
PT schizophrenia
XX
PS Claim 22; Fig 17; 205pp; English.
XX
CC The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1
CC is a gene associated with bipolar affective disorder (BAD). HKNG1
CC polynucleotides are useful to identify compounds modulating HKNG1 gene
CC expression or HKNG1 polypeptide expression/activity. Compounds inhibiting
CC or enhancing HKNG1 gene expression or activity in individuals can then
CC be administered therapeutically to treat HKNG1-mediated disorders,
CC especially neuropsychiatric disorders e.g. BAD, schizophrenia, or
CC HKNG1-mediated myopia disorders, such as early-onset autosomal
CC dominant myopia. The polynucleotides can be used in gene therapy
CC techniques to treat such disorders. They are also useful in diagnosis
CC to identify individuals having, or at risk of developing, HKNG1-mediated
CC disorders due to mutations in the HKNG1 gene. Such mutations especially
CC result in the production of a protein with a different sequence to
CC the human full-length HKNG1 polypeptide or splice variant sequences,
CC especially the substitution of a lysine for a glutamic acid at residue
CC 202 or 184. The polynucleotides are also useful in gene mapping, to
CC produce probes or primers to identify similar sequences (e.g. mutants
CC or sequences from different species) and to produce transgenic
CC animals.
XX
SQ Sequence 466 AA;

Query Match 97.4%; Score 2464; DB 20; Length 466;
Best Local Similarity 100.0%; Pred. No. 7.4e-195;
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 MKPPLLVTIVCLLWIKDSHCAPTWKDKTAISENLKSFSEGEIDADEVKKALTGIKQMK 71
Db 1 mkppllvtivclllwkdshcaptwkdktaisenlksfsegeidadeevkkaltgikmq 60
Qy 72 IMMERKEKEHTNLMSTLKKCKREEKQKALKLNEYQHELEERLCRESLADSWGCRSCL 131
Db 61 immerkekehtnlmstlkkckreekqealkllneyqheleerlcresladswgcrscl 120
Qy 132 ENNCMRIYTTQCPSSWSSVKNKIERFFRKIYQFLPPFHEDNEKDLPISEKLEEDAQLTQM 191
Db 121 enncmriyttcqpsswssvknkierffrkikyqlfpfhednekdlpisekleedaqltqm 180

QY 192 EDVFSQLTVDVNSLFRNFVFMQOEFQDTFQSHFISDQTLTEPYFFFAFSAKEPMTKA 251
 Db 181 edvfsqitvdvnslnfrsfnvfmqgefqtqfshfisdtdltepyffafskpmtka 240
 QY 252 DLEQCWDIPNFQFPCNFSVIYESVETTKMLKAIEDLPKQDKAPDHGGLSKMLPGQ 311
 Db 241 dleqcwdipnffqlfcnfsviyesvettkmlkaiedlpkgdkapdhggliskmlpgq 300
 QY 312 DRGLCGELDONLSRCFKFHEKCKQAHLSDECDPVPALHTEDEARLVNVSNOQYGOI 371
 Db 301 drglcgeidnlsrckfhekcckqahlsedcdvpalhhteidearlvnvsngqygoi 360
 QY 372 LQWTRKHLEDTAYLVEKMRGQFGVWSELANOAPETEIIFNISQVVPRIHEGNIKQDET 431
 Db 361 lqmrkhhledtaylvekmrgfgvwselanqapeteiifnsiqvvprihegniskqdetm 420
 QY 432 MTDLISILPSSNFTLKIPLESASSENFIGYVVAKALQHFKEHFKTW 477
 Db 421 mtdlsilpssnftlkipleesaessnfigyvvakalqhfkehfkwtw 466

RESULT 3
 AAY30785
 ID AAY30785 standard; Protein; 495 AA.
 AC AAY30785;
 XX
 DT 23-NOV-1999 (first entry)
 XX
 DE Protein encoded by human HKNG1 cDNA.
 XX
 KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
 KW schizophrenia; splice variant.
 XX
 OS Homo sapiens.
 XX
 XX WO9947535-A1.
 PN
 XX 23-SEP-1999.
 XX
 XX 16-MAR-1999; 99WO-US05606.
 XX
 PR 16-MAR-1998; 98US-0078044.
 PR 05-JUN-1998; 98US-0088312.
 PR 28-OCT-1998; 98US-0106056.
 PR 22-JAN-1999; 99US-0236134.
 XX
 XX (MILL-) MILLENNIUM PHARM INC.
 PA (REGC) UNIV CALIFORNIA.
 XX
 XX Chen H, Freimer NB;
 PI
 XX WPI; 1999-562047/47.
 DR N-PSDB; AAZ10750.
 XX
 XX New HKNG1 polynucleotides useful in diagnosis and treatment of
 PT neuropsychiatric disorders, e.g. bipolar affective disorders and
 PT schizophrenia .
 PT
 XX
 XX Claim 1; Fig 1A-B; 205pp; English.
 PS
 XX
 XX The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1
 CC is a gene associated with bipolar affective disorder (BAD). HKNG1
 CC polynucleotides are useful to identify compounds modulating HKNG1 gene
 CC expression or HKNG1 polypeptide expression/activity. Compounds inhibiting
 CC or enhancing HKNG1 gene expression or activity in individuals can then
 CC be administered therapeutically to treat HKNG1-mediated disorders,
 CC especially neuropsychiatric disorders e.g. BAD, schizophrenia, or
 CC HKNG1-mediated myopia disorders, such as early-onset autosomal
 CC dominant myopia. The polynucleotides can be used in gene therapy
 CC techniques to treat such disorders. They are also useful in diagnosis
 CC to identify individuals having, or at risk of developing, HKNG1-mediated

CC disorders due to mutations in the HKNG1 gene. Such mutations especially
 CC result in the production of a protein with a different sequence to
 CC the human full-length HKNG1 polypeptide or splice variant sequences,
 CC especially the substitution of a lysine for a glutamic acid at residue
 CC 202 or 184. The polynucleotides are also useful in gene mapping, to
 CC produce probes or primers to identify similar sequences (e.g. mutants
 CC or sequences from different species) and to produce transgenic
 CC animals.
 XX
 SQ Sequence 495 AA;
 Query Match 96.5%; Score 2441; DB 20; Length 495;
 Best Local Similarity 98.9%; Pred. No. 6.4e-193;
 Matches 465; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 7 SNSGNMKPPLLVTIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGIDAEVKKALTG 66
 Db 25 nnsngnmkppllvfvicllwikdshcaptwkdktaisenlksfsevgideadeevkkaltg 84
 QY 67 IKQMKIMMERKEKEHTNLMSTLKKCREKQKALKLNEVOEHLEEEERLCRESLADSWGE 126
 Db 85 ikqmkimmerkekehtnlmstllkcreekqeaalkllnevgheleeeerlcresladshge 144
 QY 127 CRSCLENNCMRIYTCOPSWSSVKNKIETFRKIYQFLFPFHEDNEKDLPISEKLIIEEDA 186
 Db 145 crscleenncmriytcopswssvknkierfrkiyqflfpfhednekdpiSekllieeda 204
 QY 187 QLTOMEDVFSQLTVDVNSLFRNFVFMQOEFQDTFQSHFISDQTLTEPYFFFAFSAKE 246
 Db 205 qltqkdvfsqitvdvnslnfrsfnvfrkqgepdtqfshfisdtdltepyffpafske 264
 QY 247 PMTKADLEQCWDIPNFQFPCNFSVIYESVETTKMLKAIEDLPKQDKAPDHGGLISK 306
 Db 265 pmtkadlegcwidipnffqlfcnfsviyesvettkmlkaiedlpkgdkapdhggglisk 324
 QY 307 MLPQDQRGLCGELDONLSRCFKFHEKCKQAHLSDECDPVPALHTEDEARLVNVSNOQ 366
 Db 325 mlpqdqrglcgeidnlsrckfhekcckqahlsedcdvpalhhteidearlvnvsng 384
 QY 367 QYGOILQMTKRKHLEDTAYLVEKMRGQFGVWSELANOAPETEIIFNISQVVPRIHEGNIK 426
 Db 385 qyggilqmtkrkhledtaylvekmrgfgvwselanqapeteiifnsiqvvprihegnisk 444
 QY 427 QDETMTMTDLISILPSSNFTLKIPLESASSENFIGYVVAKALQHFKEHFKT 476
 Db 445 qdetmtmtsilpssnftlkipleesaessnfigyvvakalqhfkehfk 494

RESULT 4
 AAY30792
 ID AAY30792 standard; Protein; 446 AA.
 AC AAY30792;
 XX
 DT 23-NOV-1999 (first entry)
 XX
 DE Mature secreted human HKNG1 protein sequence.
 XX
 KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
 KW schizophrenia; splice variant.
 OS Homo sapiens.
 XX
 XX WO9947535-A1.
 PN
 XX 23-SEP-1999.
 XX
 XX 16-MAR-1999; 99WO-US05606.
 XX
 PR 16-MAR-1998; 98US-0078044.
 PR 05-JUN-1998; 98US-0088312.

Db 1 mkppllvivylldqdcaptgkdrtsiredpkgsfageidvdeevkalligmqmk 60
 QY 72 IMMERKEHTNLMSTLKKREEKQKALKLLNEVOHEERLCLRESLADSWGECRSL 131
 Db 61 ilmerreehsklmrltkkreekqalklmevqehleeeerlcqslmgsddecksl 120
 QY 132 ENNCRIYTTCCPSWSSVKNKIERFRKIYQFLPFHEDNEKDLPISEKLEEDAQLTQM 191
 Db 121 esdcmrfyttcqswwsmkstiervfrkiyqlfpheddekeipvgekteedvqlmqi 180
 QY 192 EDVFSQLVTDVNSLFRNSFNFRQWQEFDTFQSHFISDLDLPEYFFPAPFSEKPTKA 251
 Db 181 envfsqltdvvgfynmsfhvfkmgqgefdlafgysmtdsdnepyffpafskepakka 240
 QY 252 DLEQCWDIPNFOLFPCNFVSIVSYESVETITKMLKATIEDLPKQDKAPDHGLSKMLPGQ 311
 Db 241 hpmqswdlpsffqfncfslvsygsatvtemlkatedlskqkdsahgypssattwpvr 300
 QY 312 DRGLCGELDQNLRCFHEKQKQCAHLSEDCPDVPAHTELDIAIRLVNVSNOQYQOI 371
 Db 301 grgicgepgqnsseclgharcqkqdylwadcpavpelytkadealelvnlsnqgyavq 360
 QY 372 LQMTKRKHEDTAYLVKMRGQFGWVSELANOAPETELIENSIOVVPRIHEGNISKQDET 431
 Db 361 lqmqthhledtymekmredqfvgwtelascqpsenifsfikvvpvgvhegnfskqdek 420
 QY 432 MTDSLILPSSNFTLKIPLESAESSNFYGVVAKALQHFEKHEFTW 477
 Db 421 i-disilpssnftltipleesaessdfysmlakavqhfkefksw 465

RESULT 6

AAY30787
 ID AAY30787 standard; Protein; 466 AA.
 XX
 AC AAY30787;
 XX
 DT 23-NOV-1999 (first entry)
 XX
 DE Protein encoded by the guinea pig HKNG1 ortholog gphkng1815.
 XX
 KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
 KW schizophrenia; splice variant.
 XX
 OS Cavia cobaya.
 XX
 PN W09947535-A1.
 XX
 PD 23-SEP-1999.
 XX
 PF 16-MAR-1999; 99WO-0505606.
 XX
 PR 16-MAR-1998; 98US-0078044.
 PR 05-JUN-1998; 98US-0088312.
 PR 28-OCT-1998; 98US-0106056.
 PR 22-JAN-1999; 99US-0236134.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Chen H, Freimer NB;
 XX
 DR WPI; 1999-562047/47.
 DR N-PSDB; AAZ10755.
 XX
 PT New HKNG1 polynucleotides useful in diagnosis and treatment of
 PT neuropsychiatric disorders, e.g. bipolar affective disorders and
 PT schizophrenia
 XX
 PS Claim 1; Fig 7A-B; 205pp; English.
 XX

CC The present sequence is encoded by guinea pig HKNG1 ortholog gphkng1815.
 CC HKNG1 (Hong Kong new gene 1) is a gene associated with bipolar affective
 CC disorder (BAD). HKNG1 polynucleotides are useful to identify compounds
 CC modulating HKNG1 gene expression or HKNG1 polypeptide expression/
 CC activity. Compounds inhibiting or enhancing HKNG1 gene expression or
 CC activity in individuals can then be administered therapeutically to
 CC treat HKNG1-mediated disorders, especially neuropsychiatric disorders
 CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as
 CC early-onset autosomal dominant myopia. The polynucleotides can be used
 CC in gene therapy techniques to treat such disorders. They are also useful
 CC in diagnosis to identify individuals having, or at risk of developing,
 CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such
 CC mutations especially result in the production of a protein with a
 CC different sequence to the human full-length HKNG1 polypeptide or
 CC splice variant sequences, especially the substitution of a lysine for
 CC a glutamic acid at residue 202 or 184. The polynucleotides are also
 CC useful in gene mapping, to produce probes or primers to identify
 CC similar sequences (e.g. mutants or sequences from different species)
 CC and to produce transgenic animals.
 XX
 SQ Sequence 466 AA;

Query Match 64.8%; Score 1639; DB 20; Length 466;
 Best Local Similarity 66.7%; Pred. No. 8.9e-127;
 Matches 312; Conservative 63; Mismatches 89; Indels 4; Gaps 3;

QY 12 MKPPLLVIYVILWLKSHCAPTKWKTALISENLKSFSEVGEIDAEVVKALTIKQMK 71
 Db 1 mkppllmfpvclllwlkshcaptwkdktaisenansfseageidvdvgevkaligkqm 60
 QY 72 IMMERKEHTNLMSTLKKREEKQKALKLLNEVOHEERLCLRESLADSWGECRSL 131
 Db 61 ilmerreehsklmrltkkreekqalklmevqehleeeerlcqslmgsddecksl 120
 QY 132 ENNCRIYTTCCPSWSSVKNKIERFRKIYQFLPFHEDNEKDLPISEKLEEDAQLTQM 191
 Db 121 esdcmrfyttcqswwsmkstiervfrkiyqlfpheddekeipvgekteedvqlmqi 180
 QY 192 EDVFSQLVTDVNSLFRNSFNFRQWQEFDTFQSHFISDLDLPEYFFPAPFSEKPTKA 251
 Db 180 envfsqltdvvgfynmsfhvfkmgqgefdlafgysmtdsdnepyffpafskepakka 240
 QY 252 DLEQCWDIPNFOLFPCNFVSIVSYESVETITKMLKATIEDLPKQDKAPDHGLSKMLPGQ 311
 Db 240 daepswaienvfllcnlsfsvygsvekittlratdppkqkdsnggplskilpeq 299
 QY 312 DRGLCGELDQNLRCFHEKQKQCAHLSEDCPDVPAHTELDIAIRLVNVSNOQYQOI 371
 Db 300 drgsdglglnlsdcvnrkrqckqdyisdccpnvpeylreinealrivrnsnqgydqv 359
 QY 372 LQMTKRKHEDTAYLVKMRGQFGWVSELANOAPETELIENSIOVVPRI--HGNISKQDE 429
 Db 360 vgmqthhledtllmekmredqfvgwtelascqpsenifsfikvvpvgvhegnfskqdek 419
 QY 430 TMMDLSILPSSNFTLKIPLESAESSNFYGVVAKALQHFEKHEFTW 477
 Db 420 tvvps-sllpssnftlsspleksagnanfidhvvkvqlqhfkefktw 466

RESULT 7

AAY30788
 ID AAY30788 standard; Protein; 450 AA.
 XX
 AC AAY30788;
 XX
 DT 23-NOV-1999 (first entry)
 XX
 DE Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7b.
 XX
 KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
 KW schizophrenia; splice variant.

[illegible]

CC a glutamic acid at residue 202 or 184. The polynucleotides are also
 CC useful in gene mapping, to produce probes or primers to identify
 CC similar sequences (e.g. mutants or sequences from different species)
 CC and to produce transgenic animals.
 XX
 XX

SQ Sequence 374 AA;

Query Match 50.5%; Score 1277; DB 20; Length 374;
 Best Local Similarity 54.5%; Pred. No. 4.8e-97;
 Matches 255; Conservative 44; Mismatches 73; Indels 96; Gaps 3;

QY 12 MKPPLLVFIVCLLWKLKSHCAPTWKDKTAISENLSKSFSEGEIDADEVKALTIKQOMK 71
 DB 1 mkpllpmpvcilwkdchcpcptwkdktaisenansfseagidvgevkialigkqm 60

QY 72 IMMERKEHTNLMSTLKKCKREEKQKALKLNEVOEHLEERLCRESLADSGECRSCL 131
 DB 61 immerreehsklmktlkkckeeqkalkimnevehleeeeslcqvsldswdecrac 120

QY 132 ENNCMRIYTCOPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIIEEDAQLTQM 191
 DB 121 esncmrfdttcqpawsvkn----- 140

QY 192 EDVFSQTLVDVNSLNFNSFNRMQQEFDTQFSHFISDTDLTEPYFFPAFKEPMTKA 251
 DB 141 -----mepayra 147

QY 252 DLEQCWDIPNFFQFCNFSVIYESVSETITKMLKAIEDLPKQDKAPDHGLISKMLPGQ 311
 DB 148 daepswaipnvfqlcnlsfsvygsvekltltratedppkqdkdsngqgskilpeq 207

QY 312 DRGLCGELDONLSRCFKFHEKCKQKQAHLSDECDPVPALHTELDDEAIRLVNVSNOQYQGI 371
 DB 208 drgsdgkigqnlscdvnfrcrcqkcgqylsddcpnvpelyrelnealrlvrsnqgdyq 267

QY 372 LQMTKRKHELTAYLVKMKRGQFGVWSLANQAPETETIIFNSIOVVPRI--HEGNISKODE 429
 DB 268 vqmtqyhledtllmekmrqgfwsealayspgaedifnpvkvmaisahegnssddqd 327

QY 430 TMMTDLSILPSSNFTLKIPLEESAESSNFYGYVAKALQHFKEHFTW 477
 DB 328 tvvps-silpsnftlsspleksagnanfidhvvekvlgfhkfhktw 374

RESULT 9

AA37090
 ID AAY30790 standard; Protein; 373 AA.

AC AAY30790;

XX 23-NOV-1999 (first entry)

DE Guinea pig HKNG1 ortholog gphkngl815 splice variant gphkng7d.

XX HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
 KW schizophrenia; splice variant.

OS Cavia cobaya.

XX WO947535-A1.

PN 23-SEP-1999.

XX 16-MAR-1999; 99WO-US05606.

PR 16-MAR-1998; 98US-0078044.

PR 05-JUN-1998; 98US-0088312.

PR 28-OCT-1998; 98US-0106056.

PR 22-JAN-1999; 99US-0236134.

XX (MILL-) MILLENNIUM PHARM INC.

PA (REGC) UNIV CALIFORNIA.

XX Chen H, Freimer NB;

PI WPI; 1999-562047/47.

DR N-PSDB; AA210758.

XX New HKNG1 polynucleotides useful in diagnosis and treatment of
 PT neuropsychiatric disorders, e.g. bipolar affective disorders and
 PT schizophrenia

PS Claim 1; Fig 10A-B; 205pp; English.

XX The present sequence is encoded by a guinea pig HKNG1 ortholog
 CC gphkngl815 splice variant gphkng7d. HKNG1 (Hong Kong new gene 1)
 CC is a gene associated with bipolar affective disorder (BAD). HKNG1
 CC polynucleotides are useful to identify compounds modulating HKNG1
 CC gene expression or HKNG1 polypeptide expression/activity. Compounds
 CC inhibiting or enhancing HKNG1 gene expression or activity in
 CC individuals can then be administered therapeutically to treat
 CC HKNG1-mediated disorders, especially neuropsychiatric disorders
 CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as
 CC early-onset autosomal dominant myopia. The polynucleotides can be used
 CC in gene therapy techniques to treat such disorders. They are also useful
 CC in diagnosis to identify individuals having, or at risk of developing,
 CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such
 CC mutations especially result in the production of a protein with a
 CC different sequence to the human full-length HKNG1 polypeptide or
 CC splice variant sequences, especially the substitution of a lysine for
 CC a glutamic acid at residue 202 or 184. The polynucleotides are also
 CC useful in gene mapping, to produce probes or primers to identify
 CC similar sequences (e.g. mutants or sequences from different species)
 CC and to produce transgenic animals.

XX Sequence 373 AA;

Query Match 50.3%; Score 1272.5; DB 20; Length 373;
 Best Local Similarity 54.5%; Pred. No. 1.1e-96;
 Matches 255; Conservative 45; Mismatches 71; Indels 97; Gaps 4;

QY 12 MKPPLLVFIVCLLWKLKSHCAPTWKDKTAISENLSKSFSEGEIDADEVKALTIKQOMK 71

DB 1 mkpllpmpvcilwkdchcpcptwkdktaisenansfseagidvgevkialigkqm 60

QY 72 IMMERKEHTNLMSTLKKCKREEKQKALKLNEVOEHLEERLCRESLADSGECRSCL 131

DB 61 immerreehsklmktlkkckeeqkalkimnevehleeeeslcqvsldswdecrac 120

QY 132 ENNCMRIYTCOPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIIEEDAQLTQM 191

DB 121 esncmrfdttcqpawsvkn----- 140

QY 192 EDVFSQTLVDVNSLNFNSFNRMQQEFDTQFSHFISDTDLTEPYFFPAFKEPMTKA 251

DB 141 -----mepayra 146

QY 252 DLEQCWDIPNFFQFCNFSVIYESVSETITKMLKAIEDLPKQDKAPDHGLISKMLPGQ 311

DB 147 daepswaipnvfqlcnlsfsvygsvekltltratedppkqdkdsngqgskilpeq 206

QY 312 DRGLCGELDONLSRCFKFHEKCKQKQAHLSDECDPVPALHTELDDEAIRLVNVSNOQYQGI 371

DB 207 drgsdgkigqnlscdvnfrcrcqkcgqylsddcpnvpelyrelnealrlvrsnqgdyq 266

QY 372 LQMTKRKHELTAYLVKMKRGQFGVWSLANQAPETETIIFNSIOVVPRI--HEGNISKODE 429

DB 267 vqmtqyhledtllmekmrqgfwsealayspgaedifnpvkvmaisahegnssddqd 326

QY 430 TMMTDLSILPSSNFTLKIPLEESAESSNFYGYVAKALQHFKEHFTW 477

DB 327 tvvps-silpsnftlsspleksagnanfidhvvekvlgfhkfhktw 373

peripheral nervous system diseases and neuropathies, such as Alzheimer's, Parkinson's disease, Huntington's disease, and amyotrophic lateral sclerosis. In addition, (I) is involved in chemotactic or chemokinetic activity, regulation of haematopoiesis and is useful for treating myeloid or lymphoid cell disorders, platelet disorders such as thrombocytopenia and for regeneration of bone, cartilage, tendon, ligament and/or nerve tissue growth, and in tissue repair, healing of burns, incisions, ulcers, for treating osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease. Furthermore, (I) is also useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, various immune deficiencies and disorders including severe combined immunodeficiency (SCID), bacterial or fungal infections, autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic reactions and conditions, such as asthma or other respiratory problems. In addition, (I) affects biorhythms or circadian cycles of rhythms, fertility, metabolism, catabolism, anabolism, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides analgesic effects or other pain reducing effects, immunoglobulin like activity and can act as an antigen in a vaccine composition to raise an immune response. AAU28020-AAU28395 represent novel human secreted protein amino acid sequences of the invention.

Sequence 449 AA;

Query Match 17.2%; Score 434; DB 22; Length 449;
Best Local Similarity 25.8%; Pred. No. 2.3e-27;
Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

QY 12 MKPPLLVIVCLWLKDSHCAPTWK-----DKTAISENLKSFSEGEIDADEEVKALT 65
Db 1 mmktllilfvgl-----twesqvlqdtvsdelqemsgskynkeignavn 51
QY 66 GIKOMKIMMERKEHTNLMTLKKCKREEKQKALKNLNEVQHEERLCRESLADSWG 125
Db 52 gvqkikltktneertkllisleeakkkedalnetresetkikelpgvcnetmalwe 111
QY 126 ECRSCLNNCMRIYT-TCQPSWSVKNKIERFFRKIYQFLFPFHEDNEKDLPISEKLEE 184
Db 112 eckpcklqtkmkyarvcrgsglvgqlleeflnqspfyfwmngdr-----idsilen 165
QY 185 DAOLTMEDV----FSQLTVDVNSLFRNFVRQMQOEDQTFQSHFIS-DTDLTEPYF 239
Db 166 drqqtthmlvmqdhfrassilidelqdrf-----ftrepqdy--hylpfslphrphf 218
QY 240 FPAFSKEPTMKADLE-QCWDIPNFFQFCNFVSIVSEVSETITKMLKALEDLPKQDKAP 298
Db 219 f--fksrivrslmpfspyepnlhmfqplenihea-----qqamdihfhsfaf 267
QY 299 DHGGLSKMLPGQDGLGELDONLSKCFKHEKCKQCAHLSEDC-----PDVPALHTEL 354
Db 268 qhpptefiregddrtvcvrlrhnstgclrmkqcdkcrellsvdcstnpsqaklrrel 327
QY 355 DEAIRLVNSQOYQGLTOMTRKHEDTAVLVEKMGQFCGWSELANOAPETIIINSTQ 414
Db 328 desigvaerltrkynellksyqwmkntssilleqlneqfnwvrlanltqgedqyylyrvt 387
QY 415 VVPRIHGNIKSODEMTMDLSILPSNSFTLKIPLESSESSNFIGYVVAKALQHF-KEH 473
Db 388 tvas-htsdsvpgsvgtrevvklfdspdtvtvtpvvevrknkpfmetvaekalqeyrkkh 446

RESULT 12

AAAR11704
ID AAR11704 standard; Protein; 448 AA.

AC AAR11704;

XX 20-JUN-1991 (first entry)

DE Cytolysis Inhibitor.

XX

KW cytolysis inhibitor; perforin; immunological effector molecule;
KW infertility.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..21
FT Protein /label= signal peptide
FT Protein 22..226
FT Protein /label= A-chain
FT Protein 227..448
FT Protein /label= B-chain
XX DE39333850-A.
XX 18-APR-1991.
XX 06-OCT-1989; 89DE-3933850.
XX 06-OCT-1989; 89DE-3933850.
XX (SCHD) SCHERING AG.
XX Tschoopp J, Jenne D;
XX WPI; 1991-118338/17.
XX DNA sequence coding for cytolysis inhibitor - is strong inhibitor
PT of terminal complement protein, eg perforin secreted by killer
PT cells
XX
XX Claim 13; Page 9; 15pp; German.
XX This cytolysis inhibitor is encoded by a 1.7kb BamHI-KpnI fragment
CC isolated from a liver-specific cDNA library. It is a blood plasma
CC component that inhibits immunological effector molecules. It is used
CC for systemic or local treatment of inflammatory or autoimmune
CC diseases mediated by complement or killer cells. It can also be
CC used for detoxification of membrane-active and cytolytic proteins
CC released by bacteria, fungi and insect venoms. The protein has a
CC further use in the treatment of infertility caused by deficiency of
CC cytolysis inhibitor. Monoclonal antibodies directed against the
CC cytolysis inhibitor are also covered by the invention. They are used
CC to isolate or quantify the natural protein in human plasma.
CC See also AAQ11501 and AAQ11502.
XX
XX Sequence 448 AA;

Query Match 17.1%; Score 432; DB 12; Length 448;
Best Local Similarity 25.8%; Pred. No. 3.3e-27;
Matches 123; Conservative 96; Mismatches 205; Indels 52; Gaps 14;

QY 16 LLVFIVCLWLKDSHCAPTWK-----DKTAISENLKSFSEGEIDADEEVKALTGIQK 69
Db 4 lllfvgl-----twesqvlqdtvsdelqemsgskynkeignavnqvk 54
QY 70 MKIMMERKEHTNLMTLKKCKREEKQKALKNLNEVQHEERLCRESLADSWGERS 129
Db 55 ikltiektneerktllisleeakkkedalnetresetkikelpgvcnetmalweeckp 114
QY 130 GLENNCMRIYT-TCQPSWSVKNKIERFFRKIYQFLFPFHEDNEKDLPISEKLEDAQL 188
Db 115 clkqtkmkyarvcrgsglvgqlleeflnqspfyfwmngdr-----idsilendrqq 168
QY 189 TQMEDV----FSQLTVDVNSLFRNFVRQMQOEDQTFQSHFIS-DTDLTEPYFFPAF 243
Db 169 thmlvmqdhfrassilidelqdrf-----ftrepqdy--hylpfslphrphf--f 219
QY 244 SKEPMTKADLE-QCWDIPNFFQFCNFVSIVSEVSETITKMLKALEDLPKQDKAPDHGG 302
Db 220 pksrivrslmpfspyepnlhmfqplenihea-----qqamdihfhsfafqpp 270


```
FT Protein      23..225  
FT /label= mature_clusterin_alpha_chain  
FT Protein     226..446  
FT /label= mature_clusterin_beta_chain  
FT Region       74..80  
FT /label= nuclear_localisation_signal  
FT Modified-site 86..88  
FT /label- potential_glycosylation_site  
FT Modified-site 103..105  
FT /label= potential_glycosylation_site  
FT Modified-site 145..147  
FT /label- potential_glycosylation_site  
FT Cleavage-site 225..226  
FT /note= "Interchain_cleavage_site  
FT /note= "leads to the creation of alpha and beta chains"  
FT Modified-site 290..292  
FT /label- potential_glycosylation_site  
FT Modified-site 316..318  
FT /label= potential_glycosylation_site  
FT Modified-site 353..355  
FT /label- potential_glycosylation_site  
FT Modified-site 373..375  
FT /label- potential_glycosylation_site  
XX  
XX WO200034469-A1.  
XX  
XX PN  
XX PD    15-JUN-2000.  
XX XX  
XX PF    10-DEC-1999;   99WO-US29262.  
XX XX  
XX PR    11-DEC-1998;   98US-O111856.  
XX XX  
XX PA    (UYNY ) UNIV NEW YORK STATE RES FOUND.  
XX PI Millis AJT;  
XX PX  
XX DR WPI; 2000-431300/37.  
XX  
XX PT Clusterin and gp38K-related peptide capable of altering cell migration  
PT useful for treating atherosclerosis, cancer and stenosis following  
PT vascular trauma or disease -  
XX  
XX PS Disclosure; Fig 1; 43pp; English.  
XX CC The present sequence is the protein sequence of porcine clusterin.  
CC Clusterin (also known as complement cytolysis inhibitor, sulfated  
CC glycoprotein-2, testosterone repressed prostate message-2, SP-40, 40 and  
Cc ApoJ) is essential for the migration of vascular smooth muscle cells  
CC (VSMC). The gene and protein can, therefore, be used to promote wound  
CC healing, angiogenesis and vasculogenesis, in the treatment of stenosis  
CC following vascular trauma or disease and to treat atherosclerosis, and  
CC antisense sequences can be used to treat cancer, as angioegenesis is  
CC vital for tumour survival.
```

PT Novel human secreted proteins and nucleic acids for diagnosing,
PT preventing and treating neurological, cardiovascular, infectious,
PT autoimmune, gastrointestinal, bone disorders, cancer, particularly
XX ovarian cancer

PS Claim 11; Page 386-387; 421pp; English.

XX AAD08191-AAD08213 represent cDNAs corresponding to 19 human secreted
CC protein genes and AAE03764-AAE03786 represent the proteins they encode.
CC AAE03787-AAE03800 represent human secreted protein fragments or variants.
CC The genes and their secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 19 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
XX secreted protein of the invention.

SQ Sequence 363 AA;

Query Match 12.7%; Score 320; DB 22; Length 363;
Best Local Similarity 22.6%; Pred. No. 4.3e-18;
Matches 86; Conservative 80; Mismatches 126; Indels 88; Gaps 12;
QY 16 LLVFIYCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAD-----BEVKKALTGIKQ 69
DB 4 LLLLVLGILL-----twengrvlgdgmvsdtelqemstegskyinreiknalkgkq 54
QY 70 MKTMMERKEHTNLMSTLKKCKEEKQEAALKLLNEVQEHLEEEERLCRESLADSNCEGRS 129
DB 55 ikllieqtneerkslltnleeeakkkkedaIndtkdsemkikasqgvcndtmmlweeckp 114
QY 130 CLENNCMRIYT-TCQPSWSVSKNKIERFFRKIYQFLFPFHEDNEKDLPISEKLIIEEDAQL 188
DB 115 clkqcmkfyarvcrstglvgqveeflnqspfyfwingdr-----idsllendrqq 168
QY 189 TQMEDVFSQITVDVNSLNFNSFNFFQMQEEDQTQSHFISDITLTPYFFPPAFSK--- 245
DB 169 thaldv-----mqdsfdrassi-----mdelfqdrfft-reaqdpfhfspsfsqr 213
QY 246 -----EPWTKADLEQCWDIDPFFQFCNFSVSVSETITKM 284
DB 214 rpfffnlkhrfarnlmpfpyqpl-----nfhdmfqpfddmhq-qgandvn 260
QY 285 LKAIEDLPKQDAPDHGLISKMLPGQDGLGCLGELDQNLsrcfKfHEKCKQCOAHLSEDC 344
DB 261 lhrphfpmfteedn-----qdqavckelrhnstgclkmkdqcekreillsdc 310
QY 345 -----PDVPALHTELDRAIRL 360
DB 311 ssnpaqqvlrqelnlnslqi 330